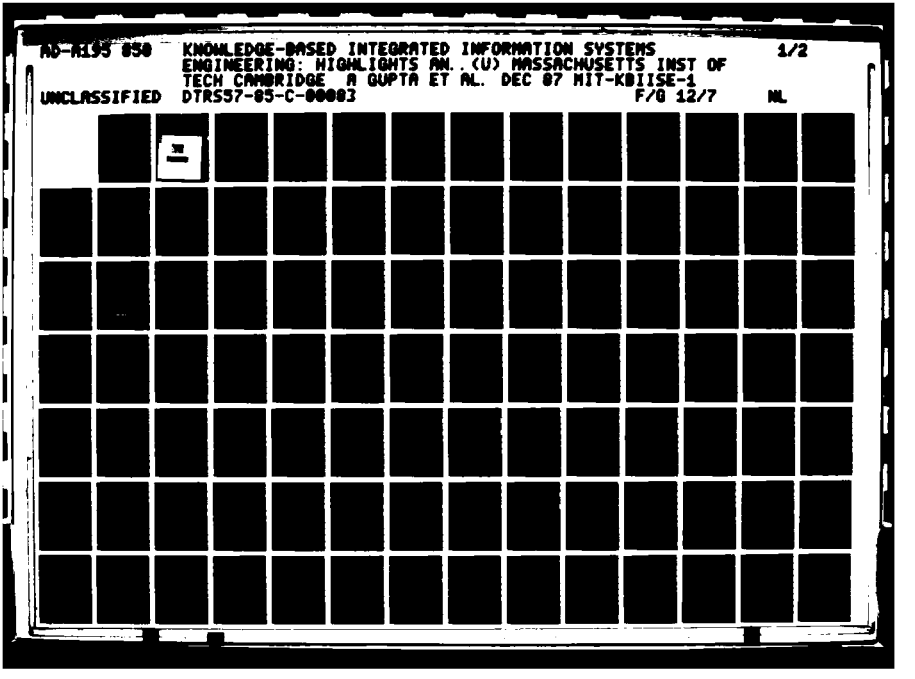
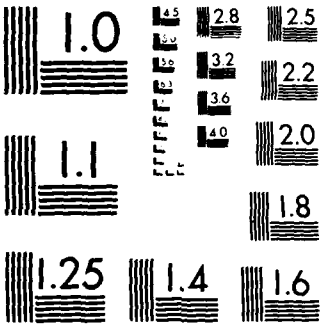
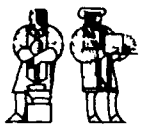

AD-A195 050
KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING: HIGHLIGHTS AN..(U) MASSACHUSETTS INST OF TECH CAMBRIDGE A GUPTA ET AL. DEC 87 MIT-KBIISE-1
1/2
UNCLASSIFIED
DTRS57-85-C-00083
F/G 12/7
NL





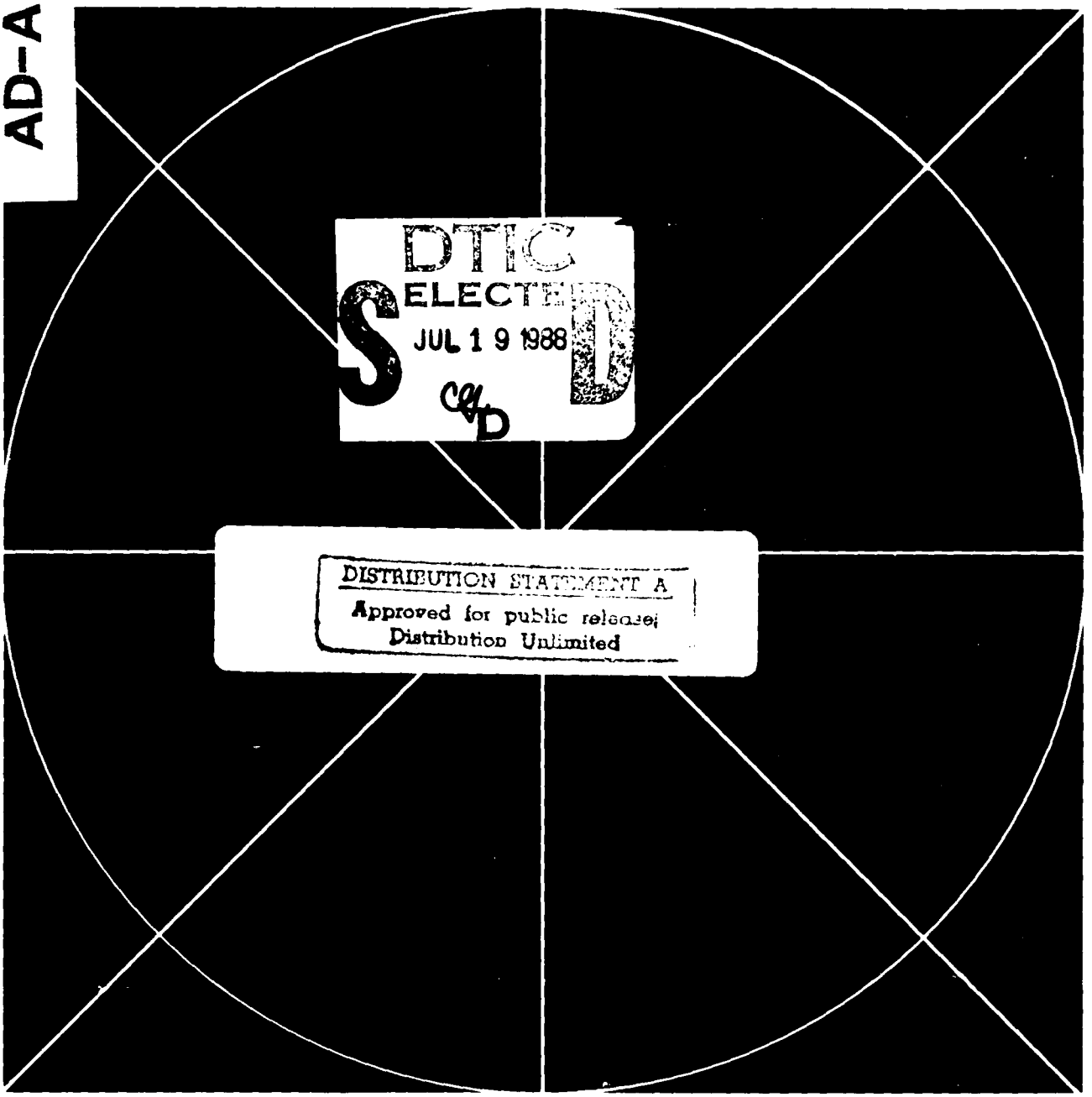
Massachusetts
Institute of
Technology

Knowledge-Based
Integrated Information
Systems Engineering
(KBIISE) Project

Volume 1

Knowledge-Based Integrated Information Systems Engineering: Highlights and Bibliography

Amar Gupta
Stuart Madnick
SERIES EDITORS



SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: MIT-KBIISE-1	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)*: Knowledge-Based Integrated Information Systems Engineering: Highlights & Bibliography	5. TYPE OF REPORT & PERIOD COVERED: Part of Final Report; Sept.86 - Jan. 88
	6. PERFORMING ORG. REPORT NUMBER: KBIISE-1
7. AUTHOR(s): Amar Gupta and Stuart Madnick (editors)	8. CONTRACT OR GRANT NUMBER(s): DTRS57-85-C-00083
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Massachusetts Institute of Technology Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Transportation Systems Center, Broadway, MA 02142 (In conjunction with U.S.A. F.)	12. REPORT DATE: December 1987 13. NUMBER OF PAGES: 129
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*: Transportation Systems Center Broadway, MA 02142	15. SECURITY CLASS. *(of this report)*: Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT *(of this Report)*: Approved for Public Release; distribution unlimited	
17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*	
18. SUPPLEMENTARY NOTES: This volume is one of eight volumes prepared by M. I. T. for Department of Transportation and Department of Defense (U.S. Air Force).	
19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*: Integration, knowledge, databases, systems engineering, methodologies, information modeling, heterogeneous database systems	
20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*: This volume presents the highlights of the Knowledge-Based Integrated Information Systems Engineering (KBIISE) project, an effort that focuses on identifying and understanding the key issues involved in the intelligent integration of large-scale distributed information systems. Underlying trends in the information technology area have led to a situation	

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

where most organizations depend on a portfolio of systems, ranging from mainframes to minicomputers, and from general purpose workstations to sophisicated CAD/CAM systems, to support their computational requirements. The large size of individual systems and the high degree of diversity between systems make the task of integrating islands of disparate information systems very formidable.

The integration of existing systems involves surmounting major technical and non-technical barriers which are described in the first part of this volume. The significance of these barriers is analyzed, and approaches for mitigating key problems are developed. Apart from the major issues of concurrency control and recovery in a heterogeneous computing environment, the technical analysis focuses on the subjects of logical connectivity and the use of expert systems technology to draw meaningful inferences in situations characterized by incomplete or conflicting information. The roles of critical success factors and interdependent value chains are examined under strategic issues. The section on organizational issues introduces concepts such as interorganizational networks and focused standards. A list of major recommendations is also presented in this part of the volume.

The second part of this volume contains a comprehensive annotated bibliography of published papers and technical reports that relate to Knowledge-Based Integrated Information Systems Engineering. This bibliography has been organized in several sections and sub-sections to facilitate direct access to the area of particular interest.

KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING : HIGHLIGHTS AND BIBLIOGRAPHY

Amar Gupta
Stuart Madnick

Series Editors

Knowledge-Based Integrated Information Systems Engineering (KBIISE) Report: Volume 1

Massachusetts Institute of Technology

This book is one of eight volumes published by M.I.T. as part of the Knowledge-Based Integrated Information Systems Engineering (KBIISE) Report. The contents of these eight volumes reflect the work performed by M.I.T. employees and students, as well as by a number of contractors and subcontractors. All requests for permission for photocopying and distribution of these volumes and individuals parts should be directed to the series' editors at M.I.T.

KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING: HIGHLIGHTS AND BIBLIOGRAPHY

About This Volume

This volume presents the highlights of the Knowledge-Based Integrated Information Systems Engineering (KBIISE) project, an effort that focuses on identifying and understanding the key issues involved in the intelligent integration of large-scale distributed information systems.

Underlying trends in the information technology area have led to a situation where most organizations depend on a portfolio of systems, ranging from mainframes to minicomputers, and from general purpose workstations to sophisticated CAD/CAM systems, to support their computational requirements. The large size of individual systems and the high degree of diversity between systems make the task of integrating islands of disparate information systems very formidable.

The integration of existing systems involves surmounting major technical and non-technical barriers which are described in the first part of this volume. The significance of these barriers is analyzed, and approaches for mitigating key problems are developed. Apart from the major issues of concurrency control and recovery in a heterogeneous computing environment, the technical analysis focuses on the subjects of logical connectivity and the use of expert systems technology to draw meaningful inferences in situations characterized by incomplete or conflicting information. The roles of critical success factors and interdependent value chains are examined under strategic issues. The section on organizational issues introduces concepts such as interorganizational networks and focused standards. A list of major recommendations is also presented in this part of the volume.

The second part of this volume contains a comprehensive annotated bibliography of published papers and technical reports that relate to Knowledge-Based Integrated Information Systems Engineering. This bibliography has been organized in several sections and sub-sections to facilitate direct access to the area of particular interest.

Table of Contents

Knowledge-Based Integrated Information System Engineering Project: Volume 1
Amar Gupta and Stuart E. Madnick, Editors

DEDICATED

TO

THE

NEXT

GENERATION

OF

PROFESSIONALS

SERIES EDITORS' NOTE

This book is one of eight volumes published by MIT as part of the Knowledge-Based Integrated Information Systems Engineering Project (KBIISE). In order to appreciate the papers in this book, it is necessary to be aware about the theme of the KBIISE project, its major objectives, and the different documents that summarize the research accomplishments to date.

Goal

The primary goal of the KBIISE project is to integrate islands of disparate information systems that characterize virtually all large organizations. The number and the size of these islands has grown over years and decades as organizations have invested in an increasing number of computer systems to support their growing reliance on computerized data. This has made the problem of integration more pronounced, complex, and challenging.

The need for multiple systems in large organizations is dictated by a combination of technical reasons (such as the desired level of processing power and the amount of storage space), organizational reasons (such as each department obtaining its own computer based on its function), and strategic reasons (such as the level of reliability, connectivity, and backup capabilities). Further, underlying trends in the information technology area have led to a situation where most organizations now depend on a portfolio of information processing machines, ranging from mainframes to minicomputers and from general purpose workstations to sophisticated CAD/CAM systems, to support their computational requirements. The tremendous diversity and the large size of the different systems make it difficult to integrate these systems.

Key Participants

The above problem is becoming increasingly evident in all large government agencies and in large development programs. In the fall of 1986, the U.S. Air Force (USAF) and the Transportation Systems Center (TSC) of the U.S. Department of Transportation approached M.I.T. to conduct and to coordinate research activity in this area in order "to develop the framework for a comprehensive methodology for large scale distributed, heterogeneous information systems which will provide: (i) the necessary structure and standards for an evolving top down global framework; (ii) simultaneous bottom up systems development; and (iii) migratory paths for existing systems."

Both USAF and TSC provided sustained assistance to members of our research team. In addition, Citibank and IBM provided some funds for research in very specific areas. One advantage of our corporate links was the opportunity to analyze and to generate case studies of actual decentralized organizational environments.

The research sponsors and MIT agreed that in order to deal with the heterogenity issue in a meaningful way, it was important that a critical mass of influential individuals participate in the development of solutions. Only through widespread discussion and acceptance of a proposed strategy would it become feasible to deal with the major problems. For these reasons, a Technical Advisory Panel (TAP) was constituted. Nominees to the TAP included experts from academic and research organizations, government agencies, computer companies, and other corporations. In addition, several subcontractors, the primary one being Texas A&M University, provided assistance in specific areas.

Technical Outputs

The scope of the work included (i) technical issues; (ii) organizational issues; and (iii) strategic issues. On the basis of exploratory research efforts in all these areas, 24 technical reports were prepared. Eighteen of these reports were generated by MIT research personnel, and their respective areas of investigation are summarized in the figure on the opposite page.

The five technical reports, not represented in the figure, are as follows:

#1. Summary.

#2. Record of discussions held at the first meeting of the Technical Advisory Panel (TAP) on February 17, 1987.

#3. Consolidated report submitted by Texas A&M University.

#21. Annotated Bibliography.

#23. Record of discussions held at the second meeting of the Technical Advisory Panel (TAP) on May 21 and 22, 1987.

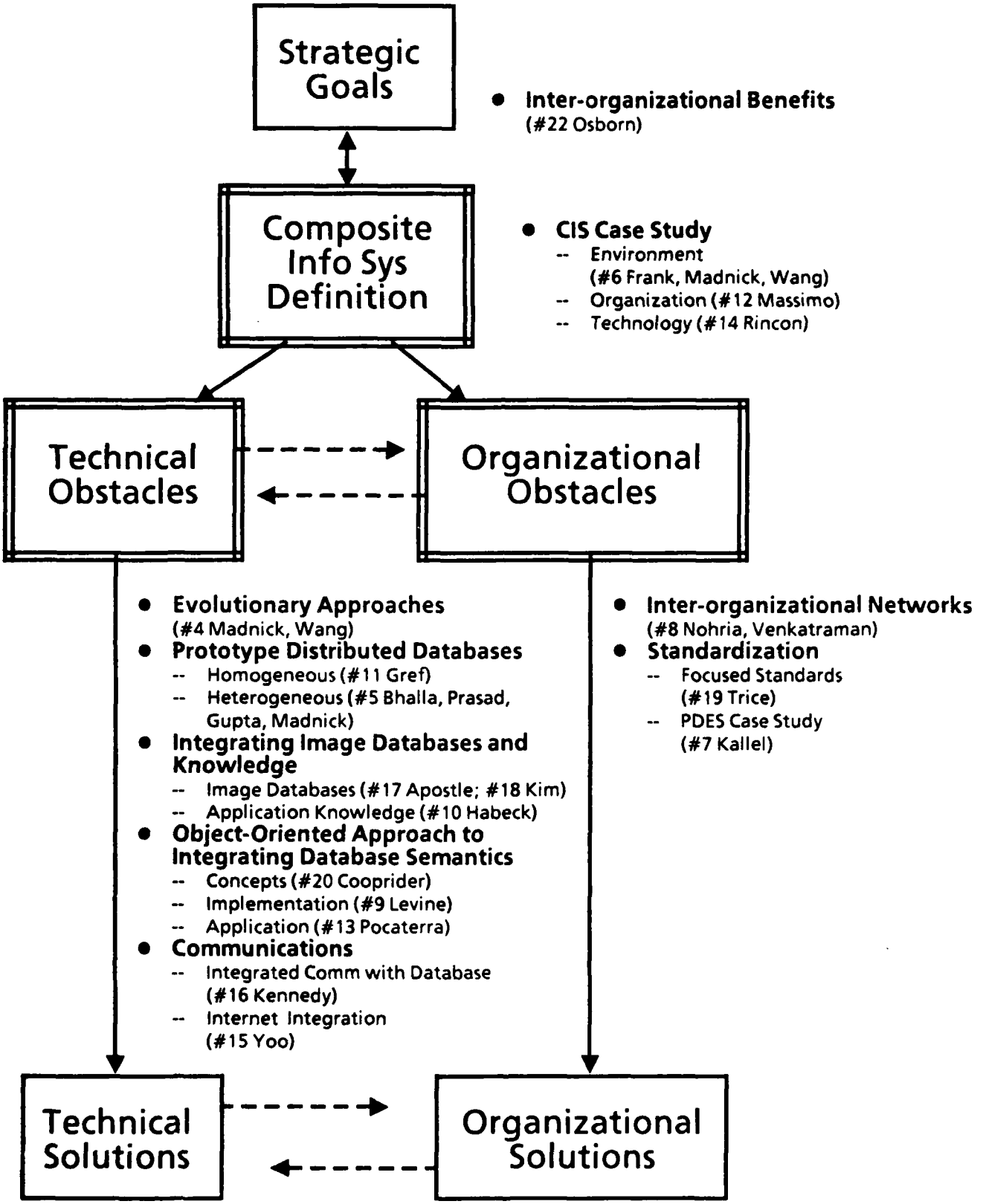
#24 Contributions received from members of the TAP highlighting their views on various aspects of the problem.

All the 24 technical reports have been edited and reorganized as an eight-volume set. The titles of the different volumes are as under:

1. KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING–HIGHLIGHTS AND BIBLIOGRAPHY
2. KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN
3. INTEGRATING DISTRIBUTED HOMOGENEOUS AND HETEROGENEOUS DATABASES – PROTOTYPES
4. OBJECT-ORIENTED APPROACH TO INTEGRATING DATABASE SEMANTICS
5. INTEGRATING IMAGES, APPLICATIONS, AND COMMUNICATIONS NETWORKS
6. STRATEGIC, ORGANIZATIONAL, AND STANDARDIZATION ASPECTS OF INTEGRATED INFORMATION SYSTEMS
7. INTEGRATING INFORMATION SYSTEMS IN A MAJOR DECENTRALIZED INTERNATIONAL ORGANIZATION
8. TECHNICAL OPINIONS REGARDING KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING

Volume 2 contains the report submitted by Texas A&M and Volume 8 highlights the views of members of the TAP. Activities described in the other 6 volumes have been conducted at MIT.

EXPLORATORY RESEARCH EFFORTS



Acknowledgments

Funds for this project have been provided by U.S. Air Force, U.S. Department of Transportation (Contract Number DTRS57-85-C-00083), IBM, and Citibank. We thank all these organizations and their representatives for their support. In particular, we are indebted to Major Paul Condit of U.S. Air Force for his initiative in sponsoring this project, to Dr. Frank Hassler, Bud Giangrande, and Bob Berk of the Transportation Systems Center (TSC) for their support and assistance, to Professor Joseph Sussman, Director, Center for Transportation Studies (CTS) at MIT for his help and encouragement, and to all the individuals whose results have been published in this book.

We would welcome receiving feedback from readers of this book.

Amar Gupta and S.E. Madnick
Massachusetts Institute of Technology
Cambridge, Massachusetts.

HIGHLIGHTS OF KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING

AMAR GUPTA AND STUART E. MADNICK

Large organizations must necessarily rely on multiple computer systems for a number of reasons, such as the increasing size of the organization and the growing reliance on computerized data. In virtually all cases, dissimilar and incompatible hardware and software systems are operating on a concurrent basis. While these systems may meet the objectives for which each was designed, their heterogeneity presents a major obstacle to ready access and assimilation of the information they contain.

The objective of Integrated Information Systems, or Composite Information Systems (CIS), is to mitigate the problem described above. Such systems must be geared to span applications, functional areas, organizational boundaries, and geographic separations in order to present a unified picture to the user. While designing such systems, it is necessary to look at a number of inter-related strategic, technical, and organizational issues. The goal of the Knowledge-Based Integrated Information Systems Engineering (KBIISE) effort, highlighted in this report, is to survey the state-of-the-art of methodologies for addressing these needs and identify areas needing increased research focus.

Strategic issues include motivating cooperation between multiple organizations, each with its own goals, priorities, and security needs. One critical success factor for such cooperation is participant consensus on the issue of access to each others' technical and non-technical information. There is an urgent need to clearly define the domains of shared information, the potential benefit to each group that participates, and the role and the responsibility of each constituent.

Under technical issues, the evolution of distributed heterogeneous information systems is studied. Being inherently more complex than conventional databases, such systems require powerful semantics and update capabilities, as well as sophisticated concurrency control and recovery mechanisms. Both the semantics and the syntax of individual queries and updates must be mapped across systems. In addition to physical connectivity issues, it becomes essential to develop new techniques for incorporating logical connectivity across systems. Such techniques combine ideas from the fields of database technology, communication technology and expert systems technology.

Organizational issues cover the process of making controlled changes in complex organizational environments. The prevailing theory of inter-organizational networks explains the multiple forces that modulate behavior of individuals, groups, and organizations. There is a need to develop focused standards to serve as the foundations for neutral representation as well as for the development of more elaborate standards.

The problem of distributed heterogeneous information systems occurs in many disciplines ranging from manufacturing to banking and from maintenance to logistics. A number of government organizations including NASA, NBS, and different agencies of the Department of Defense are faced with this problem. Within the U.S. Air Force, several efforts are directed at finding solutions in this area. In order to attain success, it is necessary that government, industry, and academia work together to develop common standards and to fill major voids that exist. A plan for concerted action is developed in this report.

SECTION 1: INTRODUCTION

Large organizations must necessarily rely on multiple computer systems to support their operations. The need for multiple systems is dictated by:

- Level of required processing power and storage space;
- Desired level of reliability and fault-tolerance;
- Geographic distribution of data collection, data manipulation, and data retrieval sites;
- Decentralized and functional structure of the organization; and
- Need to deal with different types of information, each of which may favor use of particular classes of computer hardware and software facilities.

Most organizations today depend on a portfolio of information processing machines, ranging from mainframes to microcomputers and from general purpose workstations to sophisticated CAD/CAM systems, to meet their computational requirements. While the individual hardware and software components in this growing array may meet the objectives for which each was initially designed, their heterogeneity presents a major obstacle to ready access and assimilation of information. It requires tremendous amounts of time and effort to retrieve information from multiple systems today, and in many situations, it is simply not feasible to obtain integrated responses in desired timeframes.

1.1 EXAMPLE SCENARIOS

The above problem can be better understood by looking at two different scenarios which have been identified in earlier studies done for the Air Force :

Scenario I [25]

"The air force wishes to examine the feasibility of making a modification to an aircraft. This involves changing wing materials from metal to composites. After examining the impact on the design, production, cost, and operations, the study should address life cycle cost, production feasibility and schedule for 100 new aircraft. The study should also address auxiliary issues such as retrofitting of existing aircrafts, and areas requiring further engineering, testing, and research. In order to attempt to conduct such a study, it is necessary to access engineering drawings (currently stored on paper, microfiche, CADAM, CV, and other systems), design requirements (currently stored on paper), modeling tools (such as FIFEM, NASTRAN, and other Finite Element Analysis models), data on external loads, cost histories, and information on production capacities and schedules. Today, the required information is scattered over multiple systems and several types of media. As such, the feasibility study alone may take as long as 40,000 person-hours of effort,

and it may require several months, or even years, to complete the study. The ability to intelligently access and retrieve information can significantly reduce the level of effort as well as the time delay."

Scenario II [26]

"A B-1B bomber flying in the European sector develops an operational problem. Although it is able to land safely at its base, it is determined that the problem is potentially serious. The overseas base uses electronic communication links to inform Dyess Air Force base, which in turn alerts the 15th Air Force, SAC Headquarters, Tinker Air Force Base, and the Rockwell International Support Control Center. Rockwell in turn contacts one or more of its subcontractors depending on the nature of the problem. Each of these subcontractors generates and stores technical data in digital format. However, currently the air force receives weapon system technical information on paper and microfilm. The B-1B bomber continues to remain grounded until the relevant pieces of technical information can be identified, retrieved, consolidated and studied. Only then can a combined Engineering/Manufacturing/Logistics effort be established to develop a field repair procedure."

It is difficult to put a price tag on the opportunity cost implicit in the delay in repairing the aircraft. This cost will be higher in war time than at peace time. It is clear that an accelerated pace of aircraft repair involves quick and efficient retrieval of information stored in different computers at geographically dispersed sites.

1.2 DIVERSITY OF SITUATIONS

It should be emphasized here that in both the scenarios described above, the problem is not one involving data alone. Instead, process or program related information must also be selectively retrieved to generate meaningful results. Present day systems are totally inadequate to handle either of the scenarios described above. Newer techniques must be developed to allow easy, efficient, and intelligent access to information hosted on multiple heterogeneous systems.

The various situations requiring such integration include the need to:

- *Span applications:* such as integrate CAD/CAM and documentation systems;
- *Span functional areas:* such as integrate procurement, engineering and logistics;

- *Span organizational boundaries:* such as integrate project status information from Rockwell, TRW and Lockheed; and

- *Span geographic separations:* such as integrate warehouse stockpile information from systems in USA, Europe, and Asia.

A number of specific situations have been explored by various agencies of the U.S. Government. For example, NBS has conducted research for developing automated factories of the future, and NASA has looked at the integration problem in connection with its Space Station program's information needs. Within the U.S. Air Force, there are several initiatives, including the Integrated Program for Aerospace-Vehicle Design (IPAD), the Integrated Computer-Aided Design Program (ICAD), and the broader Computer Aided Acquisition and Logistics Support (CALS) initiative. Each of these efforts looks at problems pertaining to a specific domain. However, underlying issues are almost identical in all cases.

1.3 EXISTING DEFICIENCIES

The problem of inefficient, incomplete, and time-consuming access to information can be traced, from a technical viewpoint, to functional deficiencies at several levels as described below:

- **Structured and Unstructured Applications**
  Conventional computer-based systems have been designed with specific applications in view. These systems are efficient for performing the originally intended application, but they are inefficient in dealing with ad hoc queries or new unanticipated applications.

- **Information versus Knowledge**
  Traditional database systems focus on retrieval of data, and on performing elementary operations (e.g., sorting and merging). Such systems cannot use both data and programs to respond intelligently to queries like "How many bombers can be made available to strike Target T within x hours" which require analysis of many factors.

- **Diverse Types of Information**
  Different types of information (numerical, textual, graphic, pictorial, speech, and video) are referenced in very different ways. During the sixties, computers were designed to manipulate, store, and retrieve numerical information. During the seventies, the focus was on textual information. Very little effort has so far been directed towards efficient storage and retrieval of pictorial information and its combination with numerical and textual information.

- **Communications**
  In spite of continuing efforts by a number of national and international institutions, it is still difficult to transfer complex information across computers of dissimilar architectures.

- **Granularity**
  It is difficult to judge what volume of information should be made available, and how it should be arranged and tailored to meet the needs of the person requesting the information.

- **Security**
  It is necessary to determine who can access the information, and what each person is permitted to modify. This aspect is especially important for defense applications.

- **Semantics**
  It is difficult to specify exactly the subset of information desired without knowing details of the system being accessed. Further, a particular piece of information may possess different connotations on various systems. Conversely, the same item (such as a particular spare part) may be specified by different numbers on various systems.

The above list describes a few of the key technical problems. In addition, there are major non-technical impediments that restrict integration. Because of the large number of issues and variables, a global solution to the problem of integrating distributed heterogeneous information systems has remained an elusive dream. Partial solutions are now being developed, however.

1.4 RECENT DEVELOPMENTS

The earliest research papers in the field of heterogeneous database systems were published nearly ten years ago. However, it is only in recent years that the problem has begun to receive the attention that it deserves. Currently, various governmental and professional bodies are taking special interest in this field. The February 1987 issue of Communications of the ACM carries a technical report on Heterogeneous Computing Environments [27], and the May 1987 Proceedings of the IEEE is a Special Issue dealing with Distributed Heterogeneous Systems [28].

In the Fall of 1986, the USAF and the Transportation Systems Center of the U.S. Department of Transportation, approached MIT to conduct and to coordinate research "to develop the framework for a comprehensive methodology for large scale distributed, heterogeneous information systems which will provide (1) the necessary

structure and standards for an evolving top down global framework, (2) simultaneous bottom up systems development and (3) migratory paths for existing systems." In addition to MIT personnel, several subcontractors, the primary one being Texas A&M University, were appointed to assist in specific areas, and to come up with "a structured approach to strategic planning, tactical planning, design, and implementation with methods, practices, and tools, and underlying neutral representation."

The research sponsors and MIT agreed that in order to deal with the heterogenity issue in a meaningful way, it was important that a critical mass of influential individuals agree with the suggested solution or set of solutions. Only through widespread acceptance of a proposed strategy would it become possible to deal with the major problem at hand. In response to this need, a Technical Advisory Panel (TAP) was constituted. Nominees to the TAP included experts from academic and research organizations, government agencies, government contractors, computer companies, and other corporations. The deliberations at the two TAP meetings are contained in separate technical reports [2, 23]. In all, more than two dozen leading authorities have provided their inputs through the medium of the TAP. As a core constituency of interested individuals and organizations, the TAP has been very effective both as a mechanism for suggesting new ideas, as well as a sounding board for testing innovative concepts. A list of all TAP members is enclosed in Appendix A.

The MIT research team, the subcontractors, and the TAP members studied all relevant technical, organizational, and strategic issues. After identifying the scope of the problem, a number of exploratory research efforts were initiated to assess the current state-of-the-research in the most important areas and to identify those areas and directions where further research effort is critical. These efforts culminated in the generation of 24 technical reports which have been edited and reorganized as an eight-volume set of books. The contents of each of these volumes is listed in Appendix C. Two of these volumes have been largely contributed by subcontractors and members of the TAP [3, 24]. The remaining six volumes describe the work performed at MIT [4-22]. The names of all contributors, apart from TAP members, appear in Appendix B, along with the name and the number of the Technical Report each contributor has been involved with.

The present report (Technical Report #1) describes the major findings and recommendations. Section 2 describes the overall framework and the three major thrust areas. Approaches for mitigating problems in these thrust areas are described in subsequent sections of this report.

SECTION 2: SATORI FRAMEWORK

This section describes the framework for clearly defining, articulating, and resolving the issues involved in deploying information technology within and across organizational boundaries to support strategic applications. The framework is called SATORI, which is an acronym for Strategic Applications, Technology, and Organizational Research Initiative [4,6,7]. The word SATORI connotes "a state of enlightenment or insight," according to Webster's dictionary.

2.1 KEY DIMENSIONS

The SATORI framework stresses that there are three prerequisites for attaining success. These prerequisites are: a careful determination of strategic applications and an intelligent choice of technologies, especially in relation to the organizational structure (see Figure 2.1). The absence of even one of these factors is likely to lead to failure. Strategic goals can be achieved only if appropriate technology is available; technological infrastructure must co-exist and be compatible with the organizational structure; and an organization can only exist if there was a strategy to put it in place.

The process of linking strategic goals, technical issues, and organizational aspects can be depicted as shown in Figure 2.2. At the top level, strategic goals are specified based on visions, experiences, and theories. At the next level, the Composite Information System (CIS) serves as the intermediary between the three major forces. Based on the strategic goals, technical and organizational obstacles can be identified and alternative solutions formulated. Each of the functional blocks shown in Figure 2.2 has been studied, and the numbers indicated in that figure refer to the corresponding technical reports. (These technical reports have been consolidated together in the eight volumes of this series).

An overview of all the major areas is presented in the following sections. The discussion begins with strategic issues, followed by technical issues and organizational issues. Additional information on these aspects can be found in [4, 6].

Figure 2.1

Strategic Applications, Technology, and Organization Research Initiative (SATORI)

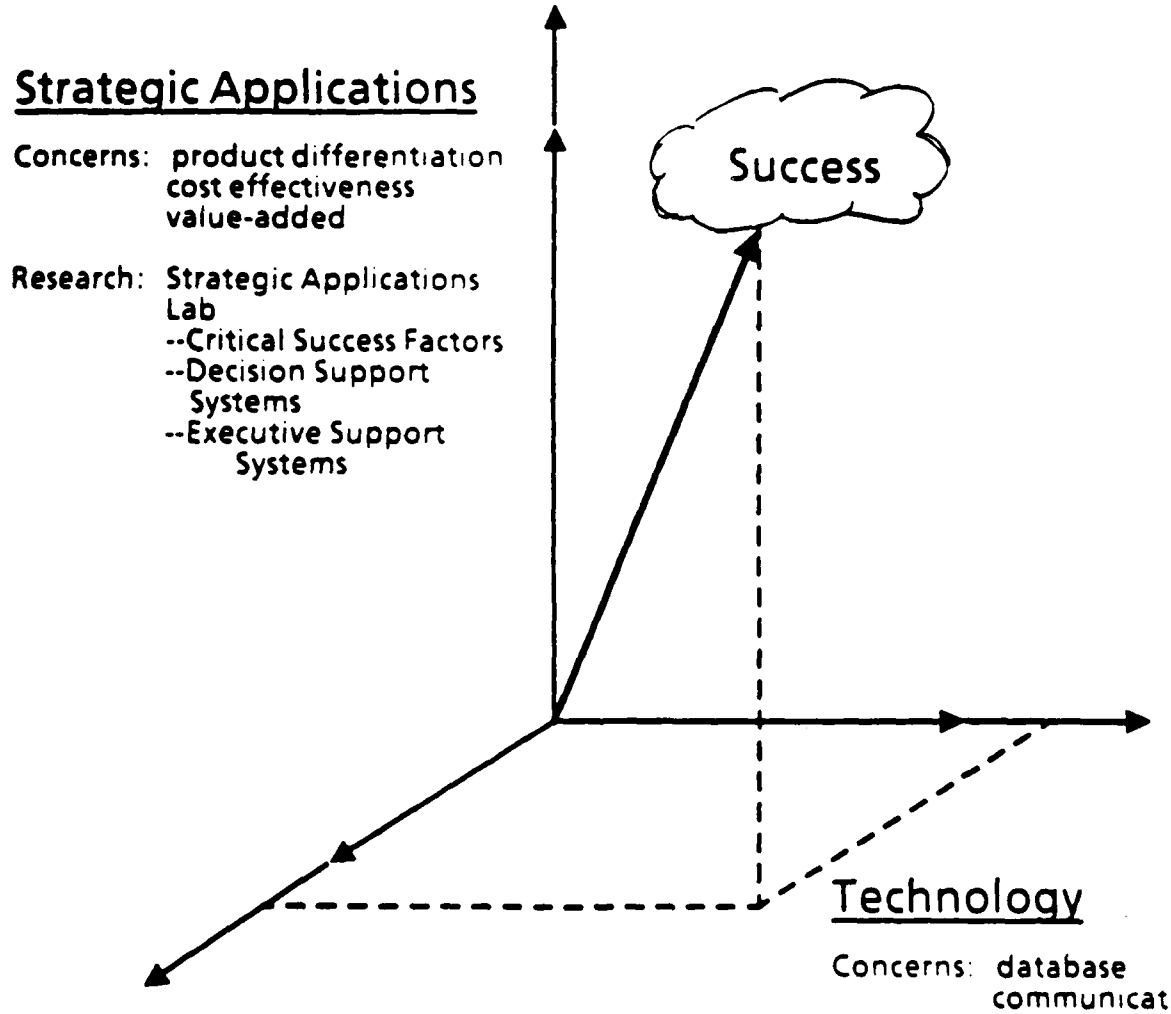
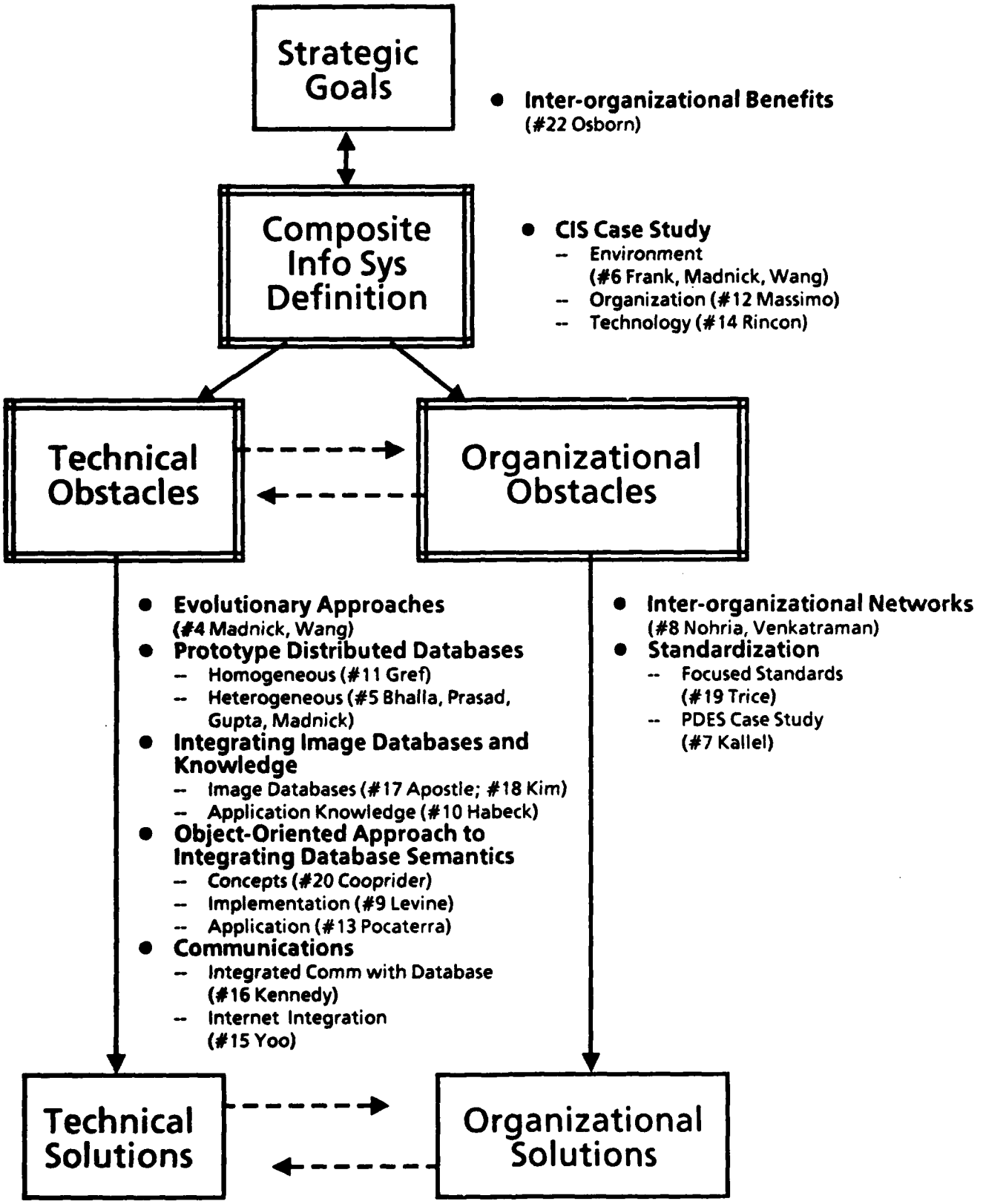


Figure 2.2

EXPLORATORY RESEARCH EFFORTS



2.2 STRATEGIC CONSIDERATIONS

The problems encountered in integrating separate systems are often expressed in terms of incompatibility between different computer systems, different operating systems, different programming languages, different database systems, different data formats, and different subsystem protocols. Presumably, these problems would not exist if one could:

- Forecast future needs perfectly;
- Communicate perfectly with all groups;
- Agree upon perfect standards for virtually everything; and
- Avoid developing new systems (e.g., product areas, modeling techniques, database systems, etc.).

None of the above constraints are likely to be overcome in the real world. As such, even if an ideal system could be built, it would be imperfect in terms of its ability to adapt to the realities of the evolving environment.

The strategic forces that impact the development of such systems, and the integration of such systems, are described in the following subsections.

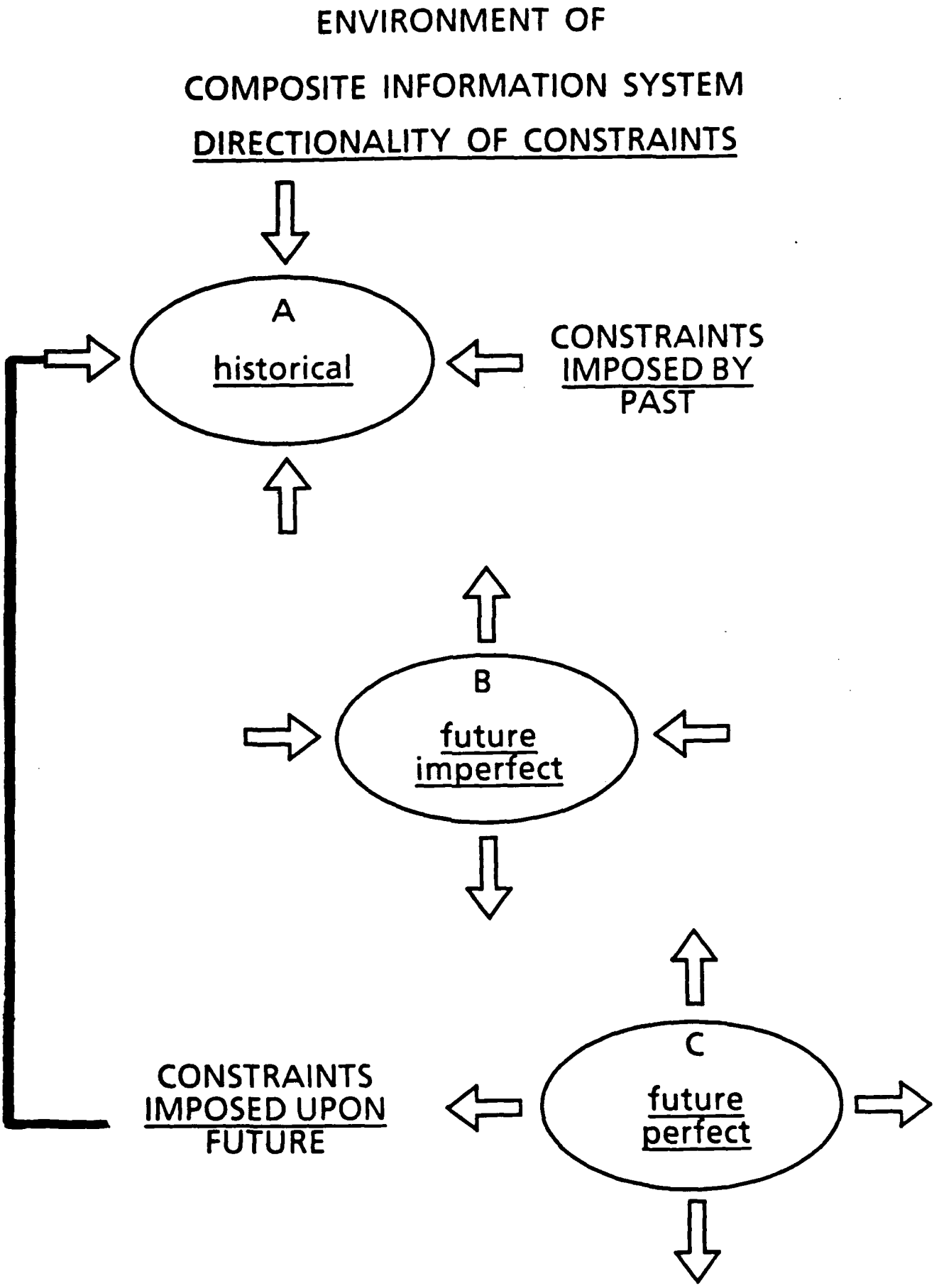
2.2.1 Directionality of Constraints

It is important to realize that there are many different environments in which a Composite Information System can be produced. These environments may require very different implementation strategies. One important way to characterize these environments is in terms of the directionality of constraints (as depicted in Figure 2.3).

The two extremes are referred to as Historical Systems and Future Perfect Systems:

- **Historical Systems** - The individual systems to be integrated already exist and are difficult, if not impossible, to change significantly. (This is sometimes referred to as the "As Is" situation). The emphasis in this situation is on the efficient, rapid and effective development of a control layer and interface transformation between the systems to enable them to communicate and act cooperatively. The constraints have been imposed by the past.

Figure 2.3


ENVIRONMENT OF
COMPOSITE INFORMATION SYSTEM
DIRECTIONALITY OF CONSTRAINTS
A
historical
CONSTRAINTS IMPOSED BY PAST
B
future imperfect
C
future perfect
CONSTRAINTS IMPOSED UPON FUTURE

- **Future Perfect Systems** - The systems to be integrated do not yet exist, thus there are no constraints imposed from the past. (This is sometimes referred to as the "To Be" situation). The emphasis in this situation is on the definition of standards, especially for data storage and communications protocols, so as to facilitate the integration of the new systems to be built.

  Over time, these systems become Historical Systems from the viewpoint of subsequent integration efforts. As such, they become constraints that will be imposed upon the future.

In reality, one is rarely dealing with either extreme. New ensembles of systems are constantly being developed (the Future Perfect situation) but must be integrated with existing systems (Historical situation). These co-existing systems are referred to as Future Imperfect Systems:

- **Future Imperfect Systems** - The systems to be integrated are both new systems under development and existing systems. Constraints on such systems are imposed by decisions made in the past, present, and future.

In order to be effective in developing integrated information systems in a Future Imperfect environment (which is the most likely case), it is important to make use of the tools and techniques from both the Historical and Future Perfect environments, and to assess the relative significance of each environment to determine emphasis. The goal is a pragmatic approach that leads to an integrated information system design and implementation in a short period of time [4].

2.2.2 Emphasis on Process

It was mentioned in Section 1.3 that computer-based systems have been conventionally designed with specific applications in view. As such they are product-oriented and not data-oriented.

Let us examine one difficulty faced in implementing the Air Force's Integrated Design Support (IDS) system prototype. The different subcontractors have been reluctant to allow the IDS prototype system developed by Rockwell to access their respective databases. Even Rockwell itself has not yet permitted its own IDS team to use production data files on an on-line basis [29]. This scenario emphasizes two issues. First, companies are reluctant to tamper with their existing product-oriented systems. Second, companies are reluctant to open up their information assets to other companies. If a company permits free access to others, it is effectively giving up its competitive edge. For further discussion of this issue, please refer to [8].

Government procurement policies usually place emphasis on end-products, and stress manufacturing drawings and data primarily to support those end-products. If the intent is to allow free access to potential competitors, the company is effectively selling the government a whole manufacturing process, inclusive of all research, design, and development information. In such a process-oriented environment, the government will possess the ability to mandate that all companies participating in the effort subscribe to the policy of full and free accessability. Such a mandate may be difficult in the present product-oriented approach, because of various commercial and legal ramifications.

2.2.3 Motivating Consensus and Cooperation

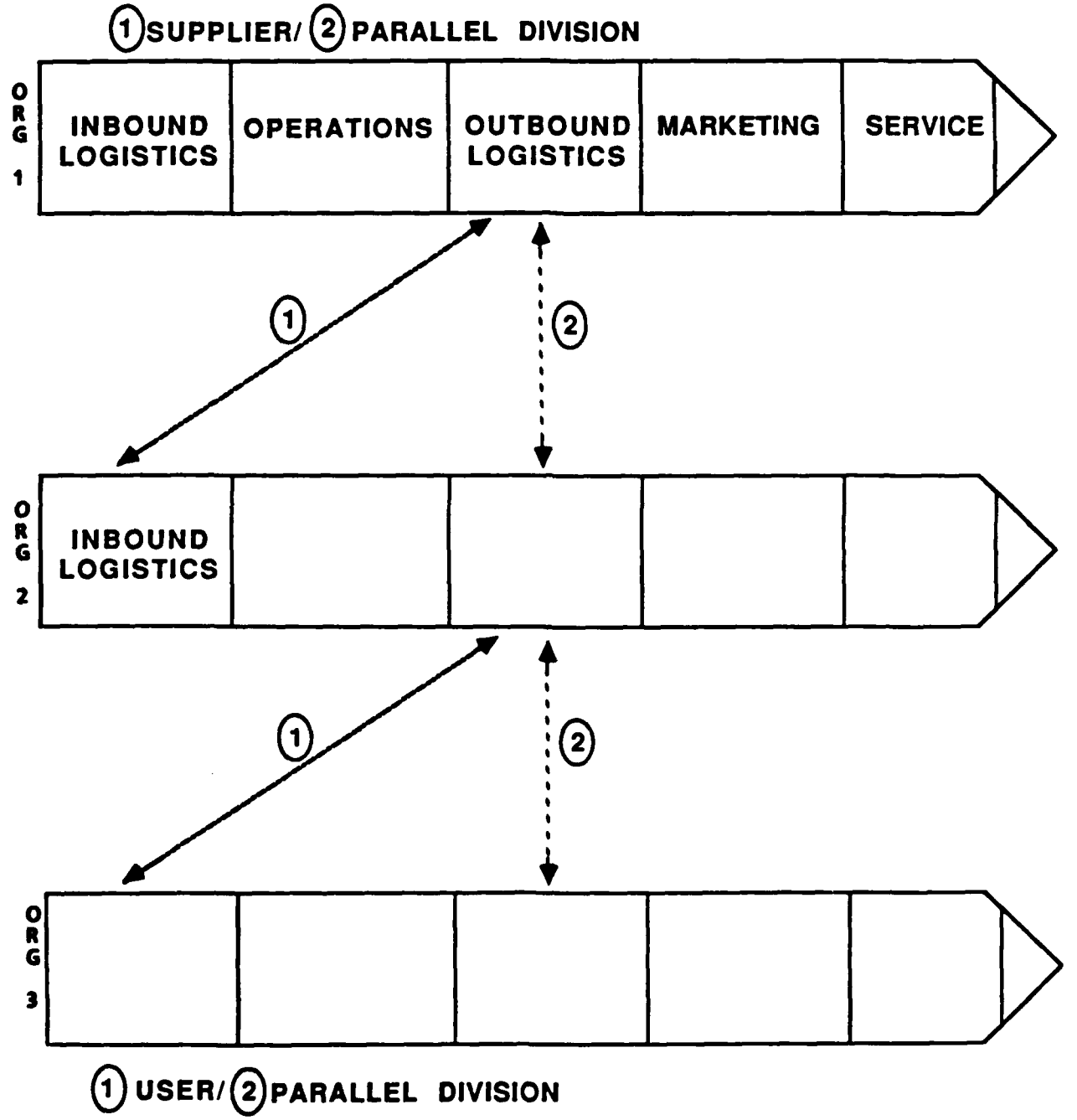
In order to achieve success in situations involving multiple groups of people, whether in the same overall organization [referred to as intra-organizational, e.g., Air Force Systems Command and Air Force Logistics Command] or in separate organizations [referred to as inter-organizational, e.g., Rockwell and TRW], it is necessary to examine the dynamics of the interactions and to think of mechanisms for motivating cooperation between all the participants and for enhancing the value to each participant.

Most large projects involve multiple organizations, loosely-coupled together as depicted in Figure 2.4. The coupling may be hierarchical (multiple layers of supplier-buyer relationships, shown with arrows marked 1) or peer-to-peer (parallel divisions of a conglomerate, shown with arrows marked 2). The question is how to motivate such organizations to develop and to deploy integrated information systems.

The manufacture of the B-1B bomber involves hierarchical-type couplings between contractors and subcontractors. Even though it has not been feasible to undertake detailed studies of these interactions, the research results gained on a similar case deserve mention here. This case involves a large mid-western hospital and physicians affiliated to it. The two key factors that make this study interesting are that (1) the hospital cannot function without doctors and the specialist doctors in this study cannot function without a hospital -- in this regard they are dependent upon each other, and (2) neither has complete control over the other, e.g., a doctor can easily decide to affiliate with another hospital, and the hospital cannot dictate how a physician runs his or her office -- in this regard they are independent of each other. We find these types of partially dependent and partially independent relationships widely prevalent in many settings. (We refer to these settings as loosely-coupled organizations).

Figure 2.4

INTER-DEPENDENT VALUE CHAINS AMONG LOOSE-COUPLED ORGANIZATIONS



The hospital perceived two critical factors that necessitated more cooperation:

(a) Percentage of office-overhead costs, which were rising and were estimated soon to exceed 60% of revenues ; and
(b) Referrals by other physicians and hospitals.

Based on the above set of perceptions, the hospital developed an integrated information system that was characterized by:

- Bi-directional benefits: The system offered value to both hospitals and physicians by supporting direct referrals between hospitals and doctors, and by simplifying administrative procedures.

- Symbiotic payoffs: In many situations, payoffs occurred only when all participants cooperated. For example, in order to provide test results in short spans of time, it was necessary to make extensive use of electronic transmissions and for both parties to cooperate.

- Asymmetrical control: The level of control exerted by each participant in the overall system does not have to be identical. In this case, the hospital was allowed to have more control over the system in return for operating the system.

The hospital used flexible low-cost technology to develop the first prototype, which was subsequently enlarged, based on reactions by the participating groups. Initial discussions about rights, responsibilities, and benefits of each group served as the major catalyst for the success of this project. In all projects involving integration of existing information systems, it is extremely critical to discuss these aspects in detail, and to make sure that all participants can potentially reap clearly defined benefits by cooperating in the integration endeavor. Additional insight into this aspect can be gained by referring to [4, 6, 8, 22].

2.2.4 Environmental Factors

The integration of diverse information systems is greatly facilitated if individual participants are willing to subscribe to a common protocol or system. Thus, increasing the level of integration may come at the cost of eroding some of the autonomy of individual participants. Examples of this problem are documented in [6, 12, 14, 22].

Autonomy of individual systems is motivated by: (i) technical reasons, such as distribution of function or characteristic of work, or geographic location of resources; (ii) organizational reasons, such as delegation of responsibility; or (iii) strategic reasons, such as security or reliability. Integration, on the other hand, is encouraged by the need to address broader issues that encompass all the individual systems, as well as by the desirability of constructively exploiting the interdependencies between systems.

In many cases it is necessary to compromise between autonomy and integration. Certain technologies discussed below provide capabilities for increasing integration with minimal reductions in autonomy. Since this balance point may itself change and evolve over time, it becomes necessary to define system architectures that provide adequate flexibility to accommodate various levels of autonomy and integration.

In the case of the B-1B bomber project, it may be inappropriate to provide access to all technical and non-technical information of all participating organizations. For example, a subcontractor, such as IBM, is not likely to voluntarily permit the lead contractor, or government auditors, to access all its data, especially those used in unrelated projects. To avoid such situations, it is essential to clearly identify what portion of data will be accessible, who can access it, who can change or update it, and who is responsible for maintaining backup copies. These issues need to be discussed and decided at senior management levels.

For further discussion of these strategic issues, please refer to [4, 6, 8].

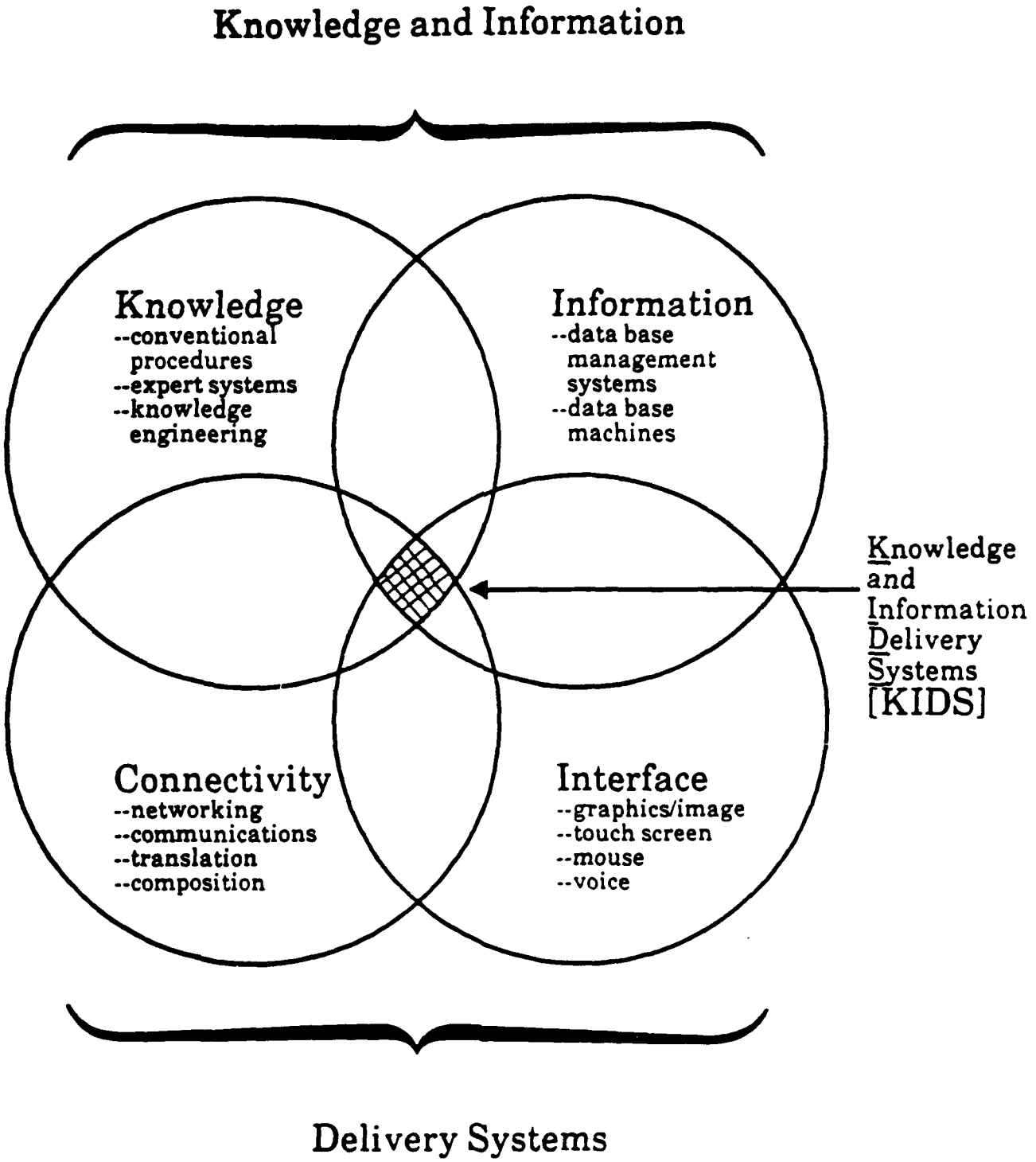
2.3 TECHNICAL CONSIDERATIONS

There are many technical obstacles that hinder the development of an integrated information system. These problems can be broadly divided into four areas: (i) information; (ii) knowledge; (iii) connectivity; and (iv) interface, as depicted in Figure 2.5. A system that addresses all these areas can be referred to as a Knowledge and Information Delivery System (KIDS). The four areas are considered in the following sections.

2.3.1 Information

Storage, manipulation, and retrieval of information in an environment characterized by diverse hardware and software involves design and use of sophisticated database management systems (DBMS). Significant voids still exist in the capabilities of available DBMS products in the context of Integrated Information Systems Engineering.

Figure 2.5


Knowledge and Information
Knowledge
--conventional procedures
--expert systems
--knowledge engineering
Information
--data base management systems
--data base machines
Knowledge and Information Delivery Systems [KIDS]
Connectivity
--networking
--communications
--translation
--composition
Interface
--graphics/image
--touch screen
--mouse
--voice
Delivery Systems

Database Management Systems were originally designed to handle large volumes of data resident on a single mainframe computer. In such centralized systems, there was no notion of parallel processing of queries by multiple computers. Further, since all information could be stored in a single format, the design of the DBMS was relatively simple.

With the growing trend towards parallel computation and storage of information on multiple computer systems, the concept of distributed database management systems evolved. For these systems, it became necessary to design mechanisms to prevent simultaneous access to data being changed by another processor. Such mechanisms, called locks, are required to prevent such concurrent access. Because of the requirement to support parallel operations in a controlled manner, distributed DBMS are much more complex than centralized DBMS, as shown in Figure 2.6.

In a distributed system, all computing elements and software can be functionally identical to each other. Such a system is characterized as a homogeneous system. The more general and complicated situation involving dissimilar computers and dissimilar DBMS software is called a heterogeneous environment. Although there are some commercially available homogeneous distributed DBMS (all with limited capabilities), heterogeneous distributed DBMS are still at an experimental stage.

Areas that require further research and development are discussed in detail in [5]. The key findings are summarized in the following paragraphs.

2.3.1.1 Homogeneous DBMS - Retrieval Only

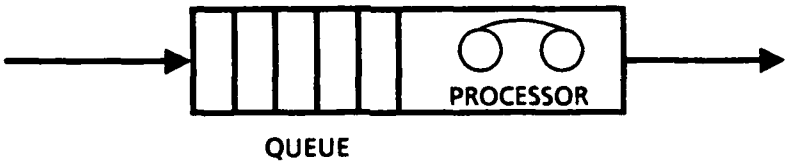
Cases of this type involve retrieval from a configuration comprised of several identical, or very similar, systems. The term "retrieval only" implies that the retrieval operation is intended in a distributed mode, while the storage operation is intended to be carried out on a "local basis" only. In such an environment, the active areas of research include:

- **Coordination of Locking Operations.** A local store operation should *not* cause an erroneous result to be generated in response to a concurrent global retrieve request.

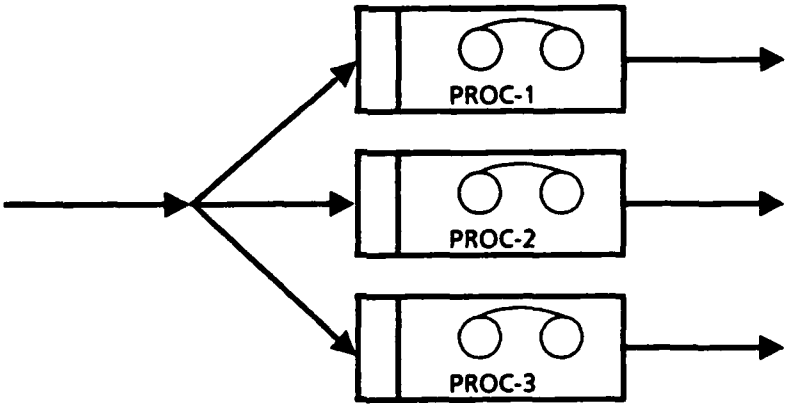
- **Optimization of Level of Pipelining and Parallelism.** This involves careful design of retrieval routines and possible segmentation of the memory space.

- **Communication Bandwidths.** The bandwidth of the communication medium imposes upper bounds on the amount of information that can be exchanged and shared in meaningful intervals of time.

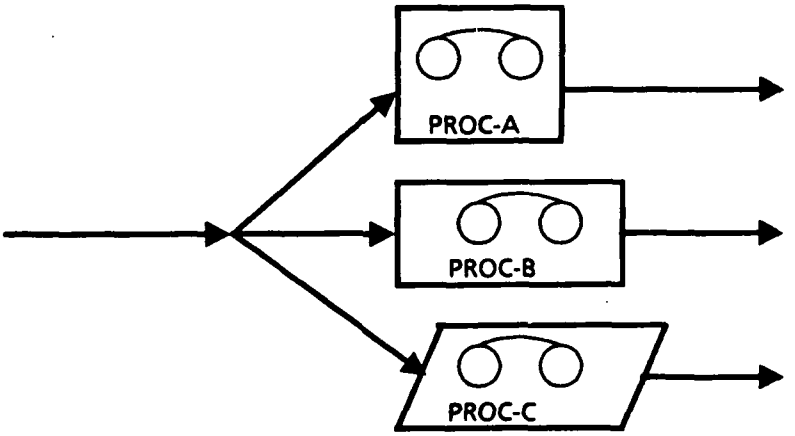
Figure 2.6



(a) Centralized Processing.
Design of DBMS is simple. Throughput is limited.



(b) Distributed Processing.
Design of DBMS is more complex. Higher throughput. All processors are identical to each other.



(c) Distributed Processing in a Heterogeneous Environment.
The design of DBMS is most complex as it must deal with idiosyncrasies of different computers.

- **Communication Protocols.** All data should not be transferred at the lowest level. Instead, high level data, models, and queries should be transmitted and deciphered by the receiving computing element.

- **Distributed Control.** No files or critical information should be maintained on a centralized basis. Also, all control and arbitration operations should be decentralized to the maximum extent possible.

2.3.1.2 Homogeneous DBMS - Retrieval and Update

In addition to the issues outlined in the previous subsection, the following issues gain increased significance:

- Multiple Copy Problem. Because of updates taking place in an asynchronous mode across different computing units, the copy of the database on one computer may be different from the database on another computer. This anomaly can lead to erroneous results.

- Disaggregation Problem. It can be difficult to decide how to split up a single update transaction to update different databases in a structured manner. For example, if a person holds five accounts in a bank, and each account is maintained on a different computer, it is difficult to know the optimal method to handle an incoming remittance - should it be split equally over the five accounts, or not?

- Concurrency Issues. Concurrent store and retrieval operations can lead to incorrect results, unless effective locking schemes are implemented.

- System Coordination Issues. If a computer generates and transmits update instructions to all other computers, what is the responsibility of the receiving station? What happens if no acknowledgement message is received? Such questions gain importance in situations where each computer belongs to different functional units which enjoy relative autonomy.

Two examples of Distributed Homogeneous Data Base Management Systems, Distributed Oracle and Distributed Ingres, have been studied in detail in [11]. The major characteristics of these two example systems are summarized in Figure 2.7.

Figure 2.7

DISTRIBUTED HOMOGENEOUS DATABASE SYSTEMS

- CASE STUDY: ORACLE AND INGRES
- TRANSPARENCY -- ACT AS A SINGLE DATABASE

PROPERTIES:	ORACLE	INGRES
1. RETRIEVAL TRANSPARENCY (Same results)	Yes	Yes
2. UPDATE TRANSPARENCY (Update from any site)	Yes	Yes
3. SCHEMA TRANSPARENCY (Schema change visible at all sites)	No (Local schema)	Yes* (Central schema)
4. PERFORMANCE TRANSPARENCY (Same from any site)	No	Yes* (Global Optimizer)
5. TRANSACTION TRANSPARENCY (Multi-update transaction done correctly)	No	No
6. COPY TRANSPARENCY (Redundant copies maintained and used efficiently)	No	No

* NOTE: INGRES' USE OF CENTRAL SCHEMA MANAGEMENT DOES POSE RELIABILITY RISKS

2.3.1.3 Heterogeneous DBMS

The above problems acquire added complexity in the case of heterogeneous environments. Since each system operates in an independent manner, it is difficult to define common standards, protocols, and etiquettes for updating information. Query requests must be converted automatically to the format and data element names appropriate for the various participating systems. Furthermore, data values retrieved must be converted to the desired format (especially if they are to be merged with values from other systems).

Most prototype systems that have been developed and refined to date, such as Multibase, have avoided the problem of global updates by not allowing global updates to take place. Some new systems are coming up with partial answers, and their salient characteristics are shown in Figure 2.8. For further details on all these systems, please refer to [5].

We now turn our attention to the more challenging task of integrating knowledge and information.

2.3.2 Knowledge Integration

The challenge involved in integrating different sources of information and knowledge can be illustrated using a bombing mission scenario. Suppose it is necessary to diagnose a hardware problem of a B-1B bomber in flight created by some unanticipated events. What procedures and tests should be conducted in time by the pilots in order not to abort the mission? Certain data needed for the diagnosis may be found in the systems available to the pilots, e.g., the expected characteristics of the hardware components and partial operational data. However, it is unlikely that on-board systems will contain the knowledge necessary to identify and rectify the problem, given the level of incomplete information.

The knowledge to perform this type of processing is probably contained on ground, partially in a DBMS which describes the hardware and partially in a Knowledge Based System (KBS) which contains the accumulated expert knowledge in diagnosing unanticipated problems. Today, no integrated system exists which allows rapid ad-hoc combination of product-definition data in a DBMS with accumulated expert knowledge in a KBS.

In order to better understand the problems involved in integrating DBMS and KBS methodologies, members of the research team developed several prototype systems. Their experiences are documented in [3, 9, 10, 13, 17, 20].

Figure 2.8

DISTRIBUTED HETEROGENEOUS DATABASE SYSTEMS

- MULTIPLE DATA MODELS (DDL)
- MULTIPLE QUERY LANGUAGES (DML)

System Name	Global Data Model	Local Data Model	Semantic Mapping Capabilities	Global Update Capabilities
ADDS	Relational	Relational, Hierarchical	Field names	Being studied
IISS	IDEF	Relational, Network	?	?
IMDAS	Semantic (SAM*)	Relational	Conversions	Only Base Relations
MERMAID	Relational	Relational	Conversion + table	Single database
MRDSM	Extended Relational	Relational	equivalence dependencies	?
MULTIBASE	Functional	Hierarchical, Network	conversions	No
NDMS	Relational	Relational, Network	?	Application programs
PRECI	Canonical	Relational	Conversion	Base relations

Conversions = unit of measure

With information and knowledge integration in mind, we now turn our attention to connectivity.

2.3.3 Connectivity

Connectivity needs to be considered at both physical and logical levels. Physical connectivity refers to the process of actual communication among disparate systems. Key issues involved in physical connectivity include problems of bandwidths, security, availability, reliability, and inter-network protocol conversions. Logical connectivity refers to the process of accessing disparate systems in concert for composite answers. A major challenge in logical connectivity is to reconcile the different assumptions and perspectives embedded in the systems being integrated.

2.3.3.1 Physical Connectivity

Based on estimates for the B-1B provided by Rockwell International, it appears that current communication bandwidths being used are inadequate to handle the Technical Order (TO) application. It is important to note that in this application, textual information is predominant and there are relatively few pictures; even when pictures are included, they are simple and invariably in 2-D. Design and manufacturing information is complex, heavily image oriented, and mostly in 3-D. Technical experts at IBM and Rockwell International have indicated that 3-D images will typically require two to three orders of magnitude larger bandwidth than equivalent 2-D images. Thus, the present communication links would be inadequate to handle the communication loads involved in transferring large amounts of information between multiple heterogeneous computers.

Two additional issues deserve mention here. First, while the concept of retrieval in a heterogeneous environment has been demonstrated, under the aegis of the IDS effort, by Rockwell International, no extensive performance and throughput study has been conducted so far. It is suggested that appropriate analytic modeling and simulation tools be designed, and comprehensive studies made into the performance aspects. Second, it is possible that for a fully loaded computational environment, response times may be unduly long as compared to current expectations. Under normal interactive operations, one desires that the response time not exceed a few seconds. However, in distributed heterogeneous environments, the response may take hours. It is rarely appreciated that the information requested may have taken years to retrieve and consolidate using conventional means. Therefore, it is important to manage user expectations. This fact reinforces the urgent need for in-depth sizing and loading studies.

In general, interconnection of heterogeneous databases makes the overall system more vulnerable to breaches of security. This is a serious drawback. Sophisticated data encryption techniques and user access keys can partially overcome this drawback. However, in most cases, the factors of simplicity and security are at the expense of each other. For further discussion of this issue, as well as related issues, please refer to [15].

We now turn our attention from physical connectivity to logical connectivity.

2.3.3.2 Logical Connectivity

In order to provide composite answers, an integration information system must, at the logical level:

- Know where all the data are stored, along with the data formats, and the local system query languages;
- Decompose the query into subqueries that can be executed by local systems;
- Accumulate the results from all the subqueries;
- Reconcile differences among the results accumulated; and
- Formulate composite answers.

To highlight the above issues, we consider an example in which three travel guides (AAA, Fodor, and the Spirit of Massachusetts, which is abbreviated as MASS in the following discussion) are accessed in concert to provide composite information about the facilities of the Logan Airport Hilton hotel in Boston.

In order to know where the data are stored and the data formats, the schemata and data dictionaries need to be accessed. Assume that a distributed DBMS can decompose the query (i.e., "What are the facilities of the Logan Airport Hilton in Boston?") into subqueries that can be executed by local DBMS's and the data from all the subqueries can be accumulated, there are still a number of semantic issues to be resolved, as illustrated by the actual examples below.

- **Synonym.** The "type-of-lodging" attribute, such as hotel, motel, and inn in AAA, is referred to as "type-of-facility" in MASS.

- **Format.** Facilities are described in a textual format in AAA (e.g., "pool"), whereas a numeric code is assigned in MASS (e.g., "6" means pool).

- **Conversion.** Each guide uses a different coding system. The meaning of each amenity code must be captured and used to convert information to a common form.

- **Incompleteness.** Each travel guide provides partial information regarding facilities at Logan Airport Hilton in Boston. The information must be merged to produce as complete an answer as possible.

- **Granularity.** Fodor simply reports whether TV is available or not, whereas AAA has three categories for TV (i.e., C/TV for color TV; CATV for cable TV; and C/CATV for color cable TV). The level of granularity may lead to contradiction as illustrated below.

- **Contradiction.** AAA indicates that Logan Airport Hilton has color TV without cable, whereas MASS reports that cable TV is available. (In reality, it has color TV with paid movies and special stations such as HBO).

- **Ambiguity.** "Room rate" in different travel guides has very different meanings. Sources of difference include whether tax, breakfast, service charge, and gratuities are included or not.

- **Inconsistency.** The name and address are reported as follows:
  AAA: Logan Airport Hilton: Logan International Airport, East Boston, 02128.
  Fodor: Hilton Inn at Logan: Logan International Airport
  MASS: The Logan Airport Hilton: Logan International Airport, Boston, 02128.
  In this case the three names differ and the two guides that list the city differ.

In order to properly integrate the diverse pieces of information, it is necessary to map synonyms and to convert different data formats. The problem is complicated by the fact that certain attributes may not be defined consistently and completely in separate systems which were independently created and administered. To resolve queries which cannot be properly handled using conventional database techniques, it becomes necessary to understand the concepts underlying the data. This process enables connections to be established based on the content of the databases. This type of exercise is referred to as *concept inferencing*.

For additional details about logical connectivity, please refer to [13, 16].

We now turn our attention to issues concerning the user-interface.

2.3.4 Interface

Information and knowledge systems can be quite challenging, especially for the casual and non-technical user. When attempting to draw upon the power of multiple systems with incompatible idiosyncrasies, the challenge can be overwhelming. Friendly user interfaces must be designed to shield the user from these idiosyncrasies.

In most cases, the user interface issue is examined when a new system is designed. However, in the case of integrating heterogeneous databases, the design of the existing databases cannot be modified. In a sense, the interface layer must be retrofitted to operate on top of previously designed systems. Since the user interface acts as the mediator between the human user and a set of divergent architectures, the task of mediation is more complex than in conventional user interfaces. Usually, a user interface resides on top of a single or homogeneous set of resources. Here, the host environments cover a broad range. As such, the user interface must be consciously designed to be readily portable across dissimilar pieces of hardware and systems software. One picture-oriented strategy has been implemented at MIT. For details, please refer to [17, 18].

Conceptually, it is feasible to use an object oriented approach to integrate heterogeneous systems involving multiple DBMS and multiple KBS. Using this approach, one can encapsulate a subsystem as an object with certain attributes, define mutually agreeable protocols among objects, and represent hierarchical inheritance properties among objects. The challenge is to realize autonomy, evolution, and integration without sacrificing other criteria such as end-user productivity and system performance - two key ingredients to commercial viability.

We now turn our attention from KIDS [4-22] to organizational obstacles and solutions.

2.4 ORGANIZATIONAL CONSIDERATIONS

Managing systems across organizations is invariably more complex than managing systems within a single organization. The role, definition, configuration, and objective of inter-organizational systems is subject to multiple forces and to continued negotiations between such forces.

Three key organizational issues are analyzed in the following subsections.

2.4.1 Multiplicity and Diversity of Goals

For the IDS project, Rockwell International is the lead contractor. Rockwell International is also the lead contractor for manufacturing the B-1B bomber, as well as for many other defense projects.

Rockwell International has a number of subcontractors, one of which is IBM. IBM in turn is subcontractor to many other defense contractors. At present, IBM holds exclusive control over its own database, and there are well defined control procedures before information from this database can be provided to other companies, including Rockwell International.

This organizational matrix raises issues of:

- **Control**: If a database is accessible to multiple companies and agencies, who has responsibility for maintaining it?

- **Motivation**: Why should a company make significant changes to its database design, simply to allow external agencies to access it with ease?

- **Competition**: If critical information is readily available to all competitors, the competitive edge of each company is lost. Does this imply that the government is buying a process rather than a product?

It must be acknowledged that companies have their own goals, which differ from those of other companies. In order to reduce conflict with corporate objectives (such as overall profitability), changes to existing computer systems may need to be minimized, even if the final result is sub-optimal. Finally, it may be necessary to introduce incentives to motivate companies to share their data with others and to modify the design of their databases to facilitate such sharing where appropriate.

For example, assume the government chooses a lead contractor, who in turn selects subcontractors. These subcontractors may in turn assign work to other companies. This reduces the amount of control and direction that can be exercised by a single organization. While the various of contractors are likely to cooperate and to share information which leads to improved operational efficiency, there exist serious concerns regarding the "proprietary" natu·e of the data [8]. Specifically:

- Participants in a given project (like the B-1B bomber project) must collaborate when their activities are interdependent, but must compete against each other (and in the broader marketplace) for other projects. Therefore, each is unlikely to view its participation in one project in isolation of its competitive position. A critical understanding of the relative balance between cooperative

and competitive roles is necessary for the effective design and implementation of any heterogeneous data systems that will involve multiple organizations.

- Information necessary to improve operational efficiency is also a source of competitive advantage for the players. Different organizations may consider different types of information proprietary (depending on their strategy perspectives), thereby restricting the overall design of the system. However, since not all organizations are likely to pursue similar strategies, there is a strong likelihood that some cooperation can be expected.

An assessment of the participants' relative competitive positions, and of the shared perceptions of their top management teams, is essential for establishing policies that encourage cooperation in a multi-organizational setting.

2.4.2 Interdependent Value Chains

As mentioned in Section 2.2.3, various participants may realize different benefits from an interorganizational system. Current research is equivocal on the issue of overall efficiency versus firm-level shifts in the sources of comparative advantage. The expectation, however, is that the possibility of obtaining firm-level advantage is on the decline in markets characterized by standardized communication systems among the vertically interconnected firms. Therefore, firms are more open to the issue of inter-organizational systems which span across multiple, interdependent value chains.

It is very important that the potential value to each organization be estimated in advance. While an organization may benefit from quick and easy access to the latest technical information, the role of the organization itself may be adversely affected by the change. For example, an organization involved in printing technical documents for the B-1B bomber may find that its services are no longer needed after an efficient inter-organizational information system becomes operational. Such an organization may be unwilling to cooperate in the integration venture.

2.4.3 Coordination and Standardization

The third organizational issue pertains to the assignment of the coordinating role and its implications on standardization. This issue has direct implications for the ownership of data. In the private sector, the coordinator is usually the entity that initiates the design and deployment of the inter-organizational network, while ownership of information is negotiated by the participants. In the context of the U.S. Air Force, the obvious solution is that the Air Force coordinates the network and owns the data. But the pivotal role played by the primary contractor makes it

difficult to separate the access privileges between the U.S. Air Force and the primary contractor.

From a functional viewpoint, the requirement is to establish a system that facilitates exchange of complete information between separate systems over the full life cycle of the products. This implies that the various participants must pool their information resources that relate to design, analysis, manufacture, testing, inspection, maintenance, and enhancement of products. This pooling can be meaningful only if a common standard can be defined covering all these diverse operating functions.

The above requirement has been recognized in many disciplines. Based on the need for exchanging information in CAD environments, the Initial Graphics Exchange Specification (IGES) was first published in 1980 and updated in 1983 and 1986. IGES information, such as drawings and 3-D wireframe models, is intended for human interpretation. Now, the effort is directed towards the Product Data Exchange Specification (PDES) which is aimed at specifying a complete product model with adequate information content to enable the specification to be directly interpreted by computer systems [7].

From all available evidence, it can take several years for such standards to be formally approved. In the interim, informal standards must be formulated through discussion and consensus. This process increases the responsibilities of the coordinator. Furthermore, evolution of informal standards is time consuming and expensive, and the standards thus devised are inherently subject to obsolescence.

Traditionally, two extreme approaches have been employed for establishing standards in organizations. In one approach, all standardization is done at the level of individual units. This approach permits greater autonomy of individual organizational entities; however, there is a lower degree of commonality across such entities. The other approach is to centralize all standardization activities; this alternative requires more time and effort. Researchers at MIT have identified an intermediate approach of "Focused Standards," which is based on the concept of critical success factors. This approach involves identification of critical entities, which are vital to the operation of the entire operation. Such entities are standardized on a centralized basis, whereas others are decided on a decentralized basis. Details about "Focused Standards" are contained in [19].

If one accepts the advantages of the Focused Standards approach, the coordinator must identify the critical entities, and also develop standards for them, based on discussions with all major participants. The procedure for deciding standards for non-critical entities must also be established. In either case, the effort must take cognizance of evolving formal standards [7, 16, 19].

It is suggested that the coordination function be performed on a centralized basis. One single unit should initiate the development of "Focused Standards", and these standards must be used for all projects. Specific implementation-related issues are discussed in the next section of this report.

SECTION 3: IMPLEMENTATION ISSUES

The need for integration of information systems is now widely accepted. This does not, however, mean that such integration is easy to accomplish. Even if physical integration is accomplished by establishing database access via network communication, the real problem of logical integration must still be assessed in many environments. If one defines the ultimate objective as "obtaining informative and correct results from integrated data bases," one still needs to resolve many technical and non-technical problems. Since many of these problems continue to be areas of active research, it is difficult to come up with a definite time-bound plan or methodology. Further, the evolution of a time-bound plan requires close familiarity with hardware and software systems in operation within the Air Force and at contractor sites. However, despite these constraints, the effort has been directed towards stating implementation-related observations in as specific terms as possible.

3.1 STRATEGIC PLAN

Issues requiring action at very senior levels within the government are delineated in this subsection.

3.1.1 Establishment of Standards

The process of integration of information systems is severely hampered by the non-availability of formal standards in many areas of interest. The process of establishing and approving standards usually takes a long time.

Consider the process of development of PDES. First, the task of identifying specifications must be assigned to the appropriate subcommittee with the right mix of expertise to avoid unnecessary bias, rework, and rejection. The different committees must consult with each other to reconcile differences in respective application models. Finally, development and dissemination of a new standard requires formal award of approvals and education of users. All this requires time and effort [7].

The costs involved in attending and participating in PDES meetings are borne by the respective members, or their employers. This implies that such participation does not take priority over other commitments of the members. A proposal, made at the April 1987 meeting, to increase the number of meetings was rejected because of the opportunity cost involved.

The government has so far taken a rather passive posture in such deliberations. It has established very few standards on its own, and it has provided little financial support to accelerate the pace of efforts conducted by other groups. Given that the government stands to derive very significant rewards from integration of information systems, immediate consideration should be given to the strategy of the government taking a more active role in establishing and promoting standards in areas of critical importance. Just as DOD supported the development and the use of Cobol, ADA, and LISP, its support of new standards will go a long way to encourage the industry and other users to conform to such standards.

As mentioned earlier in Section 2, the intermediate approach of "focused standards" can be employed, on a hierarchical basis, with each subordinate level in the hierarchy possessing the freedom to specify additional standards in its specific domain of concern [19].

3.1.2 Integration of Efforts

The process of integrating information systems is complex and expensive. Also, it is a process that cannot be viewed in isolation of other similar processes. For example, if Rockwell is manufacturing one weapon system for the Air Force and another weapon system for the Navy, it cannot be expected to modify all its computer systems to meet two conflicting sets of requirements. As such, it is highly desirable to develop solutions that can work across applications and services.

Within the context of the Air Force several programs have looked at the integration issues from different perspectives. Apart from IDS, some of the other relevant programs and systems are:

- Integrated Program for Aerospace-Vehicle Design (IPAD)
- Integrated Computer-Aided Manufacturing Program (ICAM)
- Integrated Computer-Aided Design Program (ICAD)
- Computer Aided Logistics System (CALS)
- Engineering Information Systems (EIS)

While the focus of each effort has been somewhat different, there is significant overlap in the overall goals to develop efficient mechanisms for integrating existing pieces of information. Similar concerns (e.g., for defining common standards) have been expressed by members of different teams.

The above-mentioned programs are at different stages in their respective lifecycles. Some of them, such as CALS, are in the growth phase, while others are at a stable or declining stage. To allow the newer programs to benefit from the learning phases of their predecessors, it is recommended that the funds and other resources allocated to

individual programs be pooled together. This would enable quicker implementation, as well as the creation of techniques that can be applied in multiple environments.

3.1.3 Strengthening of Contractual Relationships

It was mentioned in Section 2 that government contracts focus on products rather than on processes. The contractor provides information on products and data related to such products, but there is no requirement to cooperate with other contractors, or with the government, to permit on-line access to its information systems.

The government should consider the feasibility of including such cooperation as a requirement in future contracts, especially those relating to large weapon systems and those involving a large number of companies.

3.1.4 Identification of Weapon System

The task of integrating individual systems is currently being examined in the context of the B-1B bomber. The manufacture of this aircraft is scheduled to terminate in the near future.

The integration exercise is being attempted long after research, development, and engineering tasks had commenced and individual information systems had been created. There is little motivation to modify these information systems. One can appreciate the reluctance of the concerned parties to make modifications to such systems, or to permit on-line access to them.

In order to permit integration of data across all phases of the lifecycle, it is necessary to lay down integration guidelines from the initial stage of a project. This suggests that a new weapon system be used as the candidate for demonstrating and testing the advantages of integration. Such a strategy would allow the government to conduct a comprehensive cost-benefit analysis of integration- related amounts over the lifecycle of the project.

3.2 TACTICAL PLAN

After appropriate decisions at the strategic level have been made, it will be necessary to make second-tier decisions relating to architectures, models, and methodologies. These issues are discussed in the following subsections.

3.2.1 Quantum of Information

It has been assumed so far that rapid access to technical information will result in better and quicker decisions being made. While this assumption may be valid in many scenarios, it is not always true. Imagine a technician repairing a damaged airwing at a remote base. In response to a query to provide additional information about design and maintenance of the airwing, the computer comes up with hundreds of pages of technical data generated in several wind tunnel experiments. Instead of being an asset, such information is likely to be an unnecessary diversion for the field technician.

The application context and user profile can be important inputs in helping to quickly focus the information to the most relevant areas.

3.2.2 Neutral Representations

As mentioned earlier, companies will to be reluctant to grant access to their entire databases. Also, the significant level of heterogeneity across databases complicates the task of integration.

Instead of allowing total access to all information systems, it is better to think of subsets of information that need to be shared and to prescribe neutral representation formats for such information. Each company in the shared information pool should be made responsible for converting the relevant subset of its data from its own format to the neutral representation format. Further, the user interface should include the mechanism for converting data in the latter format into the format desired by the user [3, 7, 16, 19, 24].

A neutral representation serves as a common language for exchanging information between all individuals in a shared system. Presently, there is no single representation that covers all areas of interest. However, there are schemes with restricted applicability, such as MAP protocol for interconnecting industrial robots.

The concept of neutral representation is discussed further in [3].

3.2.3 Relevance of IDEF

Given the significant investment made by the Air Force in IDEF (IDEF0, IDEF1, IDEF2, and their derivatives), the possibility of using these modeling methodologies as the basis for neutral standards has been examined as part of the Phase I effort.

The chairman of the working group (Working Group #2) on "Evolution of Information Model" observed, on behalf of his group, that, "in the case of both IDEF0

and IDEF1, the concepts are neither fully understood nor fully documented. In fact, they are probably not even fully correct. Further, IDEF1 does not permit easy incorporation of temporal changes." One member of the same group commented on the great difficulty in transferring data from IDEF0 to IDEF1. Another member observed that if two individuals were directed to develop information models independently, they would come up with different IDEF models for the same set of activities; as such, instead of reconciling differences between multiple sets of activities, such models tend to introduce additional differences. "Blaming IDEF on these counts is like saying that the space shuttle is useless because it cannot fire torpedoes. IDEF methodologies were never designed for dealing with heterogeneous environments." Additional viewpoints about IDEF are documented in [3, 23].

In view of the above facts, it is necessary to develop a new set of information models which will serve the need for neutral representation standards. These new models, which can be termed as IDEF3, should meet the following requirements:

- They should view information in terms of objects encompassing data, relevant parts of program, and underlying assumptions.
- They should be independent of hardware and software architectures.
- They should permit entry of information from existing IDEF models.
- They should facilitate the task of conversion from one format to another.
- They should include temporal issues.
- They should allow information to be specified at various levels of abstraction, based on the needs of different audiences.

The area of information models is discussed further in [3, 23, 24].

3.3 OPERATIONAL PLAN

Design, development, and implementation of intelligent, integrated systems involves the following four key components:

- Resolution of strategic issues delineated in Section 3.1.
- Resolution of tactical issues described in Section 3.2.
- Solution of outstanding technical issues highlighted in Section 2.3.
- Development and implementation of actual systems.

In this subsection, the third and fourth components are discussed in terms of the system development process and voids in technology.

As part of this project, two draft models have been developed. The salient features of these models are described in the following paragraphs. Additional details are provided in [3].

3.3.1 Information System Evolution Process Model

The Information System Evolution Process model (ISEP-0) is aimed at systematically describing how to build and use Information Systems Integration shells, such as an IDS shell. This model is documented in Appendix A of [3]. The first section in this appendix contains a node tree of the ISEP-0 model augmented with the mechanisms required to perform each function. The second section of that appendix contains the actual IDEF0 model. The third section contains the glossary definition of each of the functions in the ISEP model [3].

3.3.2 Integrated Design System/System Development Process Model

The Integrated Design System/System Development Process Model (IDES/SPD-0) describes the process of assimilation, customization, and utilization of an IDS shell. This model has been developed down to the second/third level of decomposition with a detailed node index.

The above models, along with associated tools, techniques, methods, and methodologies are discussed in [3].

3.3.3 Research Areas

The technical areas requiring further research, and the suggested level of research activity, are summarized below.

- **Distributed Databases:** In spite of sustained efforts and repeated claims, distributed database management systems still maintain some degree of centralized information. Further, they follow an "all or nothing" approach, that is, access is permitted to the entire participating database or it is denied altogether. Further research is needed in many areas, as indicated in Section 2.3. Between 10-20 person-years of effort are needed to produce tangible results in the critical areas identified in [5].

- **Semantic Mapping:** Ideas from the realms of database management and artificial intelligence, especially expert systems, need to be combined to support efficient semantic mapping. No current system can deal with this problem in a generalized scenario. It will be necessary to devote about 10 person-years to extend the framework of [13] to meet the requirements of the IDS effort.

- **Physical Connectivity:** In order to interconnect independent resources, it is necessary to look first at the capacity of the interconnecting network, the implications of alternative protocols, and the general issues of performance, security, and reliability. Based on the knowledge gained from other programs, it is now appropriate to consider these issues in the context of IDS. The estimated level of effort is 4 person years.

- **Organization and Strategy:** It is clear that the various participating corporations (especially in the private sector) have different goals and strategies. It is naive to assume that they will participate and cooperate with data sharing to the extent desired to attain success. As an important component of the overall task, it is necessary to identify the best possible set of roles, responsibilities and relative positions of the key participants, along with a delineation of their rights, privileges, and liabilities in areas such as proprietary information and ownership of data. It is estimated that this effort will involve about 4 person years of effort for solving the problems identified in [8].

- **User Interface:** Although the field of user interface has received significant attention in recent years, the work needs to be redirected to fully insulate the user from the idiosyncrasies of the system. Also, the user interface must be capable of being hosted on top of many types of hardware and systems software. The research effort in this area is estimated at around 4 person years.

- **Standards for Neutral Representation:** Arriving at a useful standard is contingent upon further research in the areas defined above. Such a standard would must cover several different pieces of information. It is expected that significant progress toward such a standard would involve 3 person-years of effort, based on limited experience in related application environments [15,19,20,24].

In all, it is estimated that 40-50 person-years of research work is needed. By allocating an appropriate number of persons to this project, this research work can be completed within 2-3 years time. The amount of time required to implement actual systems is additional, and it is dependent on the number of systems to be linked, the type of linkage, the availability of neutral standards, and the availability of persons familiar with existing systems. An initial estimate of these activities is provided in [3].

SECTION 4: CONCLUSION

The field of Knowledge Based Integrated Information Systems Engineering raises a number of challenging strategic, technical, and organizational issues. These issues were outlined in Section 2 of this report, and they are discussed in detail in various technical reports which are contained in Volume 2 through Volume 8 of this series.

Successful implementation of an integration plan involves action at three levels as described below:

- **Strategic Level:**
  - The government should play a more active and direct role in the establishment of standards in the area of information interchange.
  - The government should consider pooling the resources of individual programs and systems with similar objectives with the present one.
  - Future contracts should address the issue of contractor cooperation with government and other contractors in implementing systems that enable information sharing across organizations.
  - A new weapon system should be identified as the basis for testing and demonstrating the feasibility of integrating heterogeneous information systems.

- **Tactical Level:**
  - The amount of the information to be made available to each category of users should be clearly delineated at the beginning stage.
  - Neutral representation standards should be established as early as possible.
  - New standards should be upward compatible with existing IDEF models to the maximum extent possible.

- **Operational Level**

  - The government should support activities aimed at filling voids in the level of required technologies.

  - The government should sponsor and fund development of prototype systems.

Based on the above requirements, it is suggested that the work be performed in four phases as described below:

Phase I: (1986-87)	Identify critical issues and generate preliminary recommendations. (This phase has been completed.)
Phase II: (1988)	Focus on specific DOD examples in depth. In parallel, conduct research aimed at defining new knowledge base architectures, and formulate standards through discussion and consensus.
Phase III: (1989)	Develop prototype solutions that embody the concept of neutral representation standards (formulated in Phase II).
Phase IV: (1990)	Develop fully operational prototypes.

The specific activities to be performed in Phase II are as follows:

- Select one (or possibly two) DOD related situations involving a large number of independent companies and diverse computer systems.

- Generate a consolidated list of all systems along with a list of files and sizes of data.

- Study the data elements that need to be transferred or shared across systems.

- Study semantic assumptions underlying each shared data element.

- Discuss possible candidates for common data representation standards with respect to all shared information.

- Estimate the frequency of the need for sharing information and the quantum of information that needs to be transferred with a view to come up with performance estimates.
- Conduct research in the areas of global updates and semantic mapping.
- Develop pragmatic strategies for accomplishing physical and logical connectivity.
- Identify types of database management systems that are best suited for the computational environments that are studied.
- Assist representatives of the government in making appropriate decisions at the strategic and tactical levels.

Phase I of this project has been performed by an academia/industry coalition coordinated by MIT. A similar structure can be potentially utilized for Phase II. By allocating sufficient resources and ensuring cooperation between participants and the organizations studied, it should be feasible to complete Phase II in one year's time.

The ultimate goal of the endeavor is to interconnect individual islands of information into an archipelago of federated knowledge. The endeavor involves many challenges as documented in these volumes. The earlier one starts the mission, the sooner one is likely to reach the goal!

REFERENCES AND BIBLIOGRAPHY

[1] Amar Gupta and Stuart E. Madnick, Executive Summary (Draft), circulated at the Second Meeting of the TAP, M.I.T., 1987 (Superceded by TR #1).

[2] Amar Gupta and Stuart E. Madnick, Record of Discussions Held at the First Meeting of the Technical Advisory Panel on February 17, 1987 , TR # 2, KBIISE Project, M.I.T., 1987 (in Volume 8, pp 5-25).

[3] Richard J. Mayer et al, Knowledge-Based Integrated Information Systems Development Methodologies Plan, TR # 3, KBIISE Project, M.I.T., 1987 (in Volume 2, pp 5-342).

[4] Stuart Madnick and Y. Richard Wang, Evolution Towards Strategic Applications Through Composite Information Systems, TR # 4, KBIISE Project, M.I.T., 1987 (in Volume 3, pp 5-30).

[5] Subhash Bhalla, B.E. Prasad, Amar Gupta and Stuart E. Madnick, A Technical Comparison of Distributed Heterogeneous Database Management Systems TR # 5, KBIISE Project, M.I.T., 1987 (in Volume 3, pp 159-218).

[6] William Frank, Stuart Madnick, and Y.Richard Wang, A Conceptual Model for Integrated Autonomous Processing, TR # 6, KBIISE Project, M.I.T., 1987 (in Volume 7, pp 5-37).

[7] Maher Kallel, Standards for Data Exchange in an Integrated Environment: A Methodological Approach, TR # 7, KBIISE Project, M.I.T., 1987 (in Volume 6, pp 137-256).

[8] Nitin Nohria and N. Venkatraman, Interorganizational Information Systems via Information Technology: A Network Perspective, TR # 8, KBIISE Project, M.I.T., 1987 (in Volume 6 , pp 65-115).

[9] Samuel P. Levine, Interfacing Objects and Databases, TR # 9, KBIISE Project, M.I.T., 1987 (in Volume 4, pp 27-168).

[10] William H.B. Habeck, An Expert System for Accessing and Integrating Design Analysis Knowledge, TR # 10, KBIISE Project , M.I.T., 1987 (in Volume 5, pp 111-256).

[11] Robert W. Gref, II, Distributed Homogeneous Database Systems: A Comparison Between ORACLE and INGRES, TR # 11, KBIISE Project, M.I.T., 1987 (in Volume 3, pp31-126).

[12] Joseph L. Massimo III, Gaining Strategic Advantage Through Composite Information Systems, TR # 12, KBIISE Project, M.I.T., 1987 (in Volume 7, pp 41-145).

[13] Marina Pocaterra Silva, A Knowledge-Based System for Resolving Semantic Conflicts: A Problem of Integrating Heterogeneous Database Management Systems, TR # 13, KBIISE Project, M.I.T., 1987 (in Volume 4, pp 171- 296).

[14] Maria de las Nieves Rincon, Integrating Systems for Financial Institutions Services Using Composite Information Systems, TR # 14, KBIISE Project, M.I.T., 1987 (in Volume 7, pp 147-283).

[15] Phillip Seung-Ho Yoo, Data Communication Networks: A Comparative Evaluation, TR # 15, KBIISE Project, M.I.T., 1987 (in Volume 5, pp 259-342).

[16] Daniel Kennedy, Achieving a Single Model for Integrating Heterogeneous Databases, TR # 16, KBIISE Project, M.I.T., 1987 (in Volume 3, pp 127-158).

[17] George Apostol, Jr., Knowledge-Based Pictorial Information Systems, TR # 17, KBIISE Project M.I.T., 1987 (in Volume 5, pp 5-65).

[18] Lowell W. Kim, Storage and Retrieval of Pictorial Information in Heterogeneous Computing Systems, TR # 18, KBIISE Project, M.I.T., 1987 (in Volume 5, pp 69-107).

[19] Andrew Trice, The Use of Standard Data Definitions in Composite Information Systems, TR # 19, KBIISE Project, M.I.T., 1987, (in Volume 6 , pp 117-135).

[20] Jay Cooprider, An Actor's Role in Integrating Expert and Database Systems, TR # 20, KBIISE Project, M.I.T., 1987 (in Volume 4, pp 5-25).

[21] Ali Tobah, Large Distributed Heterogeneous Systems : Annotated Bibliography, TR # 21, KBIISE Project, M.I.T., 1987 (in Volume 1, pp 55-125)

[22] Charley Osborn, Towards a CIS Model for Strategic Applications, TR # 22, KBIISE Project, M.I.T., 1987 (in Volume 6, pp 5-64).

[23] Amar Gupta and Stuart Madnick, Record of Discussions Held at the Second Meeting of the Technical Advisory Panel on May 21 and 22, 1987, TR # 23, KBIISE Project, M.I.T., 1987 (in Volume 8, pp 27-58).

[24] Members of the TAP, Contributions by Members of the Technical Advisory Panel (TAP), TR # 24, KBIISE Project, M.I.T., 1987 (in Volume 8, pp 59-350).

[25] R. Gurfield, Scenario : Redesign of B1 Wing (COMPWING 1), Rockwell, January 25, 1985.

[26] IDS System Design, Rockwell, May 1986, pp 36-39.

[27] D. Notkin, N. Hutchinson, J. Sanislo, and M. Schwartz, Heterogeneous Computing Environments: Report on the ACM SIGOPS Workshop on Accomodating Heterogeneity. *Communications of the ACM*, February 1987, Volume 30, No. 2, pp 132-140.

[28] W. W. Chu, Ed., Special Issue on Distributed Heterogeneous Systems, *Proceedings of the IEEE*, May 1987, Volume 75, No. 5, pp 529-736.

[29] Personal discussions with Bill Putnam and others at Rockwell on April 20, 1987.

APPENDIX A

LIST OF MEMBERS OF THE TECHNICAL ADVISORY PANEL (TAP)

	APPLETON, DANIEL S.	DACOM
*	BLOOM, HOWARD	NATIONAL BUREAU OF STANDARDS
	BROWN, ROBERT R.	RAND CORPORATION
	CHEN, PETER, P.	M.I.T. / LOUISIANA STATE UNIVERSITY
**	COLEMAN, STU	PACIFIC INFORMATION MANAGEMENT
**	GAGLIARDI, UGO	HARVARD UNIVERSITY / GENERAL SYSTEMS GROUP INC.
**	HENDERSON, JOHN	M.I.T.
	HIRSCHMAN, LYNNETTE	UNISYS
*	KNAPP, ELLEN	COOPERS AND LYBRAND
**	MANOLA, FRANK	COMPUTER CORPORATION OF AMERICA
**	MITCHELL, MARY	NATIONAL BUREAU OF STANDARDS
**	NUNAMAKER, JAY	UNIVERSITY OF ARIZONA
	NUSINOW, SAM	KNOWLEDGE BASE ENGINEERING ENTERPRISE
**	PUTNAM, BILL	ROCKWELL
*	REINER, DAVID	COMPUTER CORPORATION OF AMERICA
**	SEARS, CDR. J. ALLEN	DARPA
	STANDRIDGE, CHARLES	A. PRITSKER AND ASSOCIATES
	THOMPSON, PAUL	CONTROL DATA CORPORATION
	WIEDERHOLD, GIO	STANFORD UNIVERSITY

* *Attended first meeting only*

** *Attended second meeting only*

APPENDIX B

LIST OF CONTRIBUTORS

A large number of individuals have contributed to the twenty- four technical reports generated as part of this project. The main contributors and the technical reports they have been associated with, are as follows:

GEORGE APOSTOL, JR.
KNOWLEDGE-BASED PICTORIAL INFORMATION SYSTEMS (TR #17)

GUY BAILEY
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

SUBHASH BHALLA
A TECHNICAL COMPARISON OF DISTRIBUTED HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS (TR #5)

LELAND BLANK
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

ROBERT BROWN
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

STU COLEMAN
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

JAY COOPRIDER
AN ACTOR'S ROLE IN INTEGRATING EXPERT AND DATABASE SYSTEMS (TR #20)

TOM CULLINANE
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

WILLIAM FRANK
A CONCEPTUAL MODEL FOR INTEGRATED AUTONOMOUS PROCESSING (TR #6)

PATRICIA J. FRIEL
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

ROBERT W. GREF, II
DISTRIBUTED HOMOGENEOUS DATABASE SYSTEMS: A COMPARISON BETWEEN ORACLE AND INGRES (TR #11)

AMAR GUPTA

HIGHLIGHTS OF KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING (TR # 1)

RECORD OF DISCUSSIONS HELD AT THE FIRST MEETING OF TECHNICAL ADVISORY PANEL ON FEBRUARY 17, 1987 (TR #2)

A TECHNICAL COMPARISON OF DISTRIBUTED HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS (TR #5)

RECORD OF DISCUSSIONS HELD AT THE SECOND MEETING OF THE TECHNICAL ADVISORY PANEL HELD ON MAY 21 AND 22, 1987 (TR #23)

WILLIAM H. B. HABECK

AN EXPERT SYSTEM FOR ACCESSING AND INTEGRATING DESIGN ANALYSIS KNOWLEDGE (TR #10)

MAHER KALLEL

STANDARDS FOR DATA EXCHANGE IN AN INTEGRATED ENVIRONMENT: A METHODOLOGICAL APPROACH (TR #7)

BILL KNAPPENBERGER

KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

DANIEL KENNEDY

ACHIEVING A SINGLE MODEL FOR INTEGRATING HETEROGENEOUS DATABASES (TR #16)

LOWELL W. KIM

STORAGE AND RETRIEVAL OF PICTORIAL INFORMATION IN HETEROGENEOUS COMPUTING SYSTEMS (TR #18)

SAMUEL P. LEVINE

INTERFACING OBJECTS AND DATABASES (TR #9)

MAC LIVELY

KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

STUART E. MADNICK

HIGHLIGHTS OF KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING (TR # 1)

RECORD OF DISCUSSIONS HELD AT THE FIRST MEETING OF TECHNICAL ADVISORY PANEL ON FEBRUARY 17, 1987 (TR #2)

EVOLUTION TOWARDS STRATEGIC APPLICATIONS OF VERY LARGE DATA BASES THROUGH COMPOSITE INFORMATION SYSTEMS (TR #4)

A TECHNICAL COMPARISON OF DISTRIBUTED HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS (TR #5)

A CONCEPTUAL MODEL FOR INTEGRATED AUTONOMOUS PROCESSING (TR #6)

RECORD OF DISCUSSIONS HELD AT THE SECOND MEETING OF THE TECHNICAL ADVISORY PANEL HELD ON MAY 21 AND 22, 1987 (TR #23)

JOSEPH L. MASSIMO, III
GAINING STRATEGIC ADVANTAGE THROUGH COMPOSITE INFORMATION SYSTEMS (TR #12)

PAULA S.D. MAYER
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

RICHARD J. MAYER
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

CHRIS MENZEL
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

KEN MORRISON
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

MARIA de las NIEVES RINCON
INTEGRATING SYSTEMS FOR FINANCIAL INSTITUTIONS SERVICES USING COMPOSITE INFORMATION SYSTEMS (TR # 14)

NITIN NOHRIA
INTERORGANIZATIONAL INFORMATION SYSTEMS VIA INFORMATION TECHNOLOGY: A NETWORK PERSPECTIVE (TR #8)

GREG NUSINOW
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

SAM NUSINOW
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

CHARLEY OSBORN
TOWARDS A CIS MODEL FOR STRATEGIC APPLICATIONS (TR #22)

DON PHILLIPS
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

MARINA POCATERRA SILVA
A KNOWLEDGE-BASED SYSTEM FOR RESOLVING SEMANTIC CONFLICTS: A PROBLEM OF INTEGRATING HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS (TR #13)

B.E. PRASAD
A TECHNICAL COMPARISON OF DISTRIBUTED HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS (TR #5)

TIM RAMEY
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

PHILLIP SEUNG-HO YOO
DATA COMMUNICATION NETWORKS: A COMPARATIVE EVALUATION (TR #15)

FRANKLIN SKIPPER
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

ALI TOBAH
LARGE DISTRIBUTED HETEROGENEOUS SYSTEMS: ANNOTATED BIBLIOGRAPHY (TR #21)

PAUL THOMPSON
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

ANDREW TRICE
THE USE OF STANDARD DATA DEFINITIONS IN COMPOSITE INFORMATION SYSTEMS (TR #19)

ALVIN J. UNDERBRINK, JR.
KNOWLEDGE BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN (TR #3)

N. VENKATRAMAN
INTERORGANIZATIONAL INFORMATION SYSTEMS VIA INFORMATION TECHNOLOGY: A NETWORK PERSPECTIVE (TR #8)

Y. RICHARD WANG
A CONCEPTUAL MODEL FOR INTEGRATED AUTONOMOUS PROCESSING (TR #6)

[TR # 24 HAS BEEN CONTRIBUTED BY MEMBERS OF THE TECHNICAL ADVISORY PANEL (TAP). THEIR NAMES ARE LISTED IN APPENDIX A.]

APPENDIX C

CONTENTS OF INDIVIDUAL VOLUMES

VOLUME 1:
KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING: HIGHLIGHTS AND BIBLIOGRAPHY

VOLUME 2:
KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS DEVELOPMENT METHODOLOGIES PLAN

VOLUME 3:
INTEGRATING DISTRIBUTED HOMOGENEOUS AND HETEROGENEOUS DATABASES: PROTOTYPES

VOLUME 4:
OBJECT-ORIENTED APPROACH TO INTEGRATING DATABASE SEMANTICS

TR # 20:	AN ACTOR'S ROLE IN INTEGRATING EXPERT AND DATABASE SYSTEMS	JAY COOPRIDER
TR # 9:	INTERFACING OBJECTS AND DATABASES	SAMUEL P. LEVINE
TR # 13:	A KNOWLEDGE-BASED SYSTEM FOR RESOLVING SEMANTIC CONFLICTS: A PROBLEM OF INTEGRATING HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS	MARINA POCATERRA SILVA

VOLUME 5:
INTEGRATING IMAGES, APPLICATIONS AND COMMUNICATIONS NETWORKS

TR #17:	KNOWLEDGE-BASED PICTORIAL INFORMATION SYSTEMS	GEORGE APOSTOL, JR.
TR # 18:	STORAGE AND RETRIEVAL OF PICTORIAL INFORMATION IN HETEROGENEOUS COMPUTING SYSTEMS	LOWELL W. KIM
TR #10:	AN EXPERT SYSTEM FOR ACCESSING AND INTEGRATING DESIGN ANALYSIS KNOWLEDGE	WILLIAM H.B. HABECK
TR # 15:	DATA COMMUNICATION NETWORKS: A COMPARATIVE EVALUATION	PHILLIP SEUNG-HO YOO

VOLUME 6:
STRATEGIC, ORGANIZATIONAL AND STANDARDIZATION ASPECTS OF INTEGRATED INFORMATION SYSTEMS

TR #22:	TOWARDS A CIS MODEL FOR STRATEGIC APPLICATIONS	CHARLEY OSBORN
TR # 8:	INTERORGANIZATIONAL INFORMATION SYSTEMS VIA INFORMATION TECHNOLOGY: A NETWORK PERSPECTIVE	NITIN NOHRIA AND N. VENKATRAMAN
TR # 19:	THE USE OF STANDARD DATA DEFINITIONS IN COMPOSITE INFORMATION SYSTEMS	ANDREW TRICE
TR # 7:	STANDARDS FOR DATA EXCHANGE IN AN *INTEGRATED ENVIRONMENT*: A METHODOLOGICAL APPROACH	MAHER KALLEL

VOLUME 7:
INTEGRATING INFORMATION SYSTEMS IN A MAJOR DECENTRALIZED INTERNATIONAL ORGANIZATION

TR # 6:	A CONCEPTUAL MODEL FOR INTEGRATED AUTONOMOUS PROCESSING	WILLIAM FRANK ET. AL.
TR # 12:	GAINING STRATEGIC ADVANTAGE THROUGH COMPOSITE INFORMATION SYSTEMS	JOSEPH L. MASSIMO,III
TR # 14:	INTEGRATING SYSTEMS FOR FINANCIAL INSTITUTIONS USING COMPOSITE INFORMATION SYSTEMS	MARIA DE LAS NIEVES RINCON

VOLUME 8:
TECHNICAL OPINIONS REGARDING KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING

TR # 2:	RECORD OF DISCUSSIONS HELD AT THE FIRST MEETING OF THE TECHNICAL ADVISORY PANEL ON FEBRUARY 17,1987	AMAR GUPTA AND STUART E.MADNICK
TR # 23:	RECORD OF DISCUSSIONS HELD AT THE SECOND MEETING OF THE TECHNICAL ADVISORY PANEL ON MAY 21 AND 22, 1987	AMAR GUPTA AND STUART E. MADNICK
TR # 24:	CONTRIBUTIONS BY MEMBERS OF THE TECHNICAL ADVISORY PANEL (TAP)	

LARGE DISTRIBUTED HETEROGENEOUS SYSTEMS ANNOTATED BIBLIOGRAPHY

ALI TOBAH

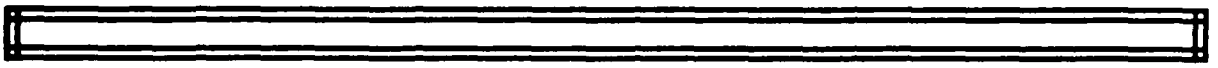
Design and integration of large distributed heterogeneous systems involves concepts from several fields. The relevant literature in these fields has been closely studied and reviewed by members of the research team as part of this project.

This technical report contains material obtained by the research team from published sources, as well as material provided or suggested by members of the Technical Advisory Panel. It corresponds to Version 6 of the Annotated Bibliography, and has been organized in three main sections, and several sub-sections, as shown below:

TECHNICAL REPORT #21

A- General Problems and Proposed Approaches.

This section presents efforts at defining the issues involved in integrating distributed, heterogeneous systems, as well as proposed approaches to the general problem. Topics range from business applications to manufacturing and engineering problems, which demonstrates the relevance of research in this area to a variety of fields. Consequently, the partial solutions and helpful concepts summarized in the following sections have roots in different disciplines.



Appleton, Daniel S.

THE TECHNOLOGY OF DATA INTEGRATION.

Datamation, November 1, 1985, pp.106-116.

The author presents software integration as the "Holy Grail of information resource management", and states that what is needed is efficient access through a single query language and database schema to data in preexisting, heterogeneous, distributed applications. He states that over the last five years the problem has been resolved into four subsets of technology: (1) user interface technology, (2) distributed heterogenous data management technology, (3) network transaction management technology and (4) interprocessor communication technology. The main focus of the article is on the second subset.

"There are two distinct concepts and approaches to distributed, heterogeneous data management . . . the interfacing approach and the integration approach. The key difference between them is that in the interfacing approach there are no technological controls over data integrity, in the integration approach, data integrity in a distributed heterogeneous environment is a major issue." He then describes the two approaches, showing the roles of the schema.

Finally, he declares the need, in order to solve the problem, for commitments to heterogeneity, to standards and to data integrity.

The article presents a brief overview of four projects attempting to solve the problem: Multibase, Integrated Information Support System (IISS), Integrated Manufacturing Distributed Database Administration System (IMDAS) and Integrated Design Support System (IDS). It also states that integration technology integrates both the current and the future integrated automated environments.

Barkmeyer, Edward; Mitchell, Mary; Mikkilineni, Krishna P.; Su, Stanley Y. W. and Lam, Herman

AN ARCHITECTURE FOR DISTRIBUTED DATA MANAGEMENT IN COMPUTER INTEGRATED MANUFACTURING.

National Bureau of Standards, NBSIR 86-3312, January 1986, 56 pgs.

The National Bureau of Standards Automated Manufacturing Research Facility (AMRF) is being developed as a testbed for automated small batch manufacturing, and will be used to support experimentation in automated metrology and factory integration. The paper presents a distributed database management system architecture to support integrated manufacturing. The system was designed to support the AMRF.

The authors discuss the requirements of the AMRF and the implications of these requirements on the management of data in a Computer Integrated Manufacturing (CIM) system. The requirements are the following:

1) integration of heterogeneous systems;
2) component systems with different data management abilities;
3) both autonomous and integrated operation of subsystems;
4) flexible manufacturing;
5) time critical operations; and
6) use of adaptive control techniques.

Then the overall design along with details of the major modules of the Integrated Manufacturing Database Administration System (IMDAS) are presented, based on the above mentioned requirements. (See also: Libes and Barkmeyer)

Dickson, Gary W.; Leitheiser, Robert L.; Wetherbe, James C. and Nechis, Mal

KEY INFORMATION SYSTEMS ISSUES FOR THE 1980's.

MIS Quarterly, September 1984, pp. 135-159.

A Delphi study, using leading IS professionals, was used to identify and rank ten key IS management issues for the 1980's. Measures were also taken of the amount of agreement achieved on these issues and rankings. The number one issue identified was "improved IS planning" followed by "facilitation and management of end user computing". The article describes the research approach involved and discusses the results.

Note that five of the ten issues would be served to some extent by integrating an organization's [hetrogeneous, distributed] systems into one system in which interfaces are transparent to the user:

- Facilitation of organizational learning and usage of IS technologies.
- Facilitation and management of end user computing.
- Improved software development and quality.
- Integration of data processing, office automation, and telecommunications.
- Measuring and improving IS effectiveness/productivity.

The first four would also be facilitated by introducing standards in the IS or computer science fields.

Three other issues agree with results obtained by Hackathorn and Karimi (1986):

- Aligning the IS organization with that of the enterprise.
- Effective use of the organization's data resources.
- Improved IS planning.

Dutton, David L.

IN PURSUIT OF CIM.

Datamation, February 1, 1986, pp. 63-66.

Computer-integrated manufacturing (CIM) is the sharing of manufacturing resources related to information collection, storage, processing and distribution in such a way as to optimize the

performance of the total enterprise. The Microelectronics Circuits Division (MCD) of Hughes Aircraft Co. is discussed as an environment that appears to need CIM. The MCD set up a CIM program whose staff determined that the highest priority in the near term would be to implement the missing elements of its manufacturing resources planning system. This led to the creation of an advanced shop floor control (SFC) system project.

There are three categories of SFC system requirements: functional, performance and advanced CIM architecture. The CIM architecture includes linkages needed to integrate the SFC system with the automatic materials ha ıdling equipment and the computer aided manufacturing line.

Some problems have surfaced, some of which are intertwined technical and management issues: Because many technical details are not yet resolved, the scope, cost and schedule are somewhat open-ended. Concern over the added complexity introduced by the integrated architecture causes division among the project sponsors. Grass-roots support for CIM projects is quite evident, and many people have called for users to manage system development projects. However, this would give the projects too narrow a support base to appreciate the proper systemic relationships. Longtime departmental barriers exist, restricting the flow of information so that the value of a common, integrated database is not readily apparent. There is also the problem of the proliferation of many different and highly customized manufacturing control systems.

Hewitt, Carl and De Jong, Peter

OPEN SYSTEMS.

In On Conceptual Modelling. Perspectives from AI, Databases and Programming Languages, Brodie, Mylopoulos and Schmidt (eds.), 1984, pp. 147-164.

Abstract: This chapter describes some problems and opportunities associated with conceptual modelling for the kind of "open systems" we foresee developing in the future. Computer applications will be based on communication between subsystems that will have been developed separately and independently. Some of the reasons for independent development are: competition, economics, geographical distribution, and diverse goals and responsibilities. We must deal with all the problems that arise from conceptual disparities. Subsystems will be open-ended and incremental -- undergoing continual evolution. There are no global objects. The only thing that all the various subsystems hold in common is the ability to communicate with each other. Message Passing Semantics is a methodology that we are developing to deal with highly parallel, distributed, open systems. Various aspects of this methodology deal with communication, description, transaction management, problem solving, change, completeness, and self-knowledge.

Madnick, S. E. and Hsu, M.

COMPOSITE INFORMATION SYSTEMS: A STRATEGY FOR SYSTEM DESIGN. (Draft, to be superseded)

Center for Information Systems Research, MIT, June 1983, 40 pgs.

The purpose of the paper is to bring about a higher level of awareness of the problem of information system integration by presenting a concept of information system development called the Composite Information System (CIS). This approach recognizes the dilemma between distributed responsibilities for system sub-units (which enhances system effectiveness and user satisfaction) and system integration. It consists of a set of principles which provide a staged path for the evolution of fragmented systems into composite information systems.

The essence of the CIS approach is captured in four principles: (1) the environment principle, which allows for multiplicity of information system components; (2) the principle for separation of data, which advocates separating a flexible data component from the processing component, thus introducing 'process descriptors' and 'database descriptors'; (3) the principle for using software tools, which provides guidelines for developing the processing components within a CIS; and (4) the principle for composition, which describes the kinds of facilities that a CIS uses to accomplish communications among component systems.

The framework for evolution towards CIS consists of four stages in increasing levels of sophistication: (1) a composite system with a mediator (or CIS executive, which uses the concept of virtual terminals to communicate with existing systems); (2) a composite system with processing partitioned from the database logically; (3) a composite system with processing partitioned from the database physically; and (4) a composite system with specialized hardware for processing, database and CIS executive.

Melkanoff, Michel A.

THE *CIMS* DATABASE: GOALS, PROBLEMS, CASE STUDIES AND PROPOSED APPROACHES OUTLINED.

Industrial Engineering, November 1984, pp. 78-93.

The author states that the goals of the ideal CIMS database are to store and provide access to the information required in the three areas of a manufacturing concern: (1) design and engineering,(2) production planning and manufacturing and (3) administration and business operations. They may be achieved by providing tools for performing common operations on data. Problems exist due to heterogeneity of the hardware, software and models as well as general DB problems. The proposed approach involves three phases of planning, implementation and evaluation including the involvement of top management. Two case studies showing partially integrated CIMS databases are the Ingersoll Milling Machine Co. and the Boeing Commercial Airplane Company.

The author states that there is no fully integrated CIMS database. Presenting the problems (see above) in creating such a DB gives a good view of the components involved.

Nusinow, Edwin I. (Sam)

CONCEPTS FOR A MANUFACTURING ORIENTED COMPUTER SYSTEM.

Centerville, Ohio, July 1983, 6 pgs.

The author first describes some of the characteristics associated with evolving computer systems that support computer integrated manufacturing. Integration is defined as a mechanism to bring together disparate elements and subsystems in order to work as a system to address some problem. Then he describes the need for common computer system functions to support product manufacturing. Next, he explores the requirements for common system manufacturing functions. In order to improve integration, system components must be well defined in terms of their interfaces, functionality and relationship to other components. A discussion on the properties of system integration follows. Four ways in which integration can be accomplished are defined: physical, logical, operational and project. Finally, some candidate functions are proposed for incorporation into a manufacturing oriented computer system architecture. Among those mentioned are project management, configuration management, production planning, document library and project glossary.

Nusinow, Sam (E. I.)

AN ENGINEERING/MANUFACTURING ENTERPRISE INTEGRATED INFORMATION CONTROL SYSTEM.

IFIP/IFAC Fifth International Conference, September 1985, 8 pgs.

Some of the integration and control problems that exist in today's engineering and manufacturing environments are discussed. They are associated with the evolution of engineering and manufacturing product data. Since understanding the enterprise is essential before the benefits of implementing computer-aided technology can be realized, enterprise analysis (top down and bottom up) techniques are explored. System interface requirements, operational business rules, product data life cycles, improvement benefits and the functional specifications needed for system implementation are identified. High level system architecture requirements are discussed, and several functions are identified: application interface, user interface, executive and data controller.

Shuey, Richard

GUEST EDITOR'S INTRODUCTION.

Computer, January 1986, pp. 14-15.

"For example, the distribution of mini and micro computers among users created an environment that, unless properly controlled, fostered independent and uncoordinated application development. Distributed computers and their associated databases created new

database requirements and constraints and introduced a host of communications-related problems and opportunities. These challenges are being addressed, but there has not been time to develop a satisfactory overall engineering discipline and an associated set of tools to handle design problems associated with data usage in distributed systems.

At the same time, the expansion in the application areas to which computers were applied (CAD/CAM and AI are examples) introduced new and more sophisticated requirements for database management systems. Thus, although there is a mature discipline for database design, new applications have created R&D opportunities that must be addressed in order to extend that existing discipline to meet current and future needs."

In meeting data requirements, data engineering must address:

- The design of databases and database management systems.
- The evolving and changing data requirements of existing applications and systems.
- The development of technology and design methods that will meet the data management needs of new application areas, such as AI. Implicit is the need to evolve from data management to the management of information and knowledge.
- The merger of relevant parts of database and communication engineering to address the data-management and control problems within large distributed-information systems.

Shuey, Richard and Wiederhold, Gio

DATA ENGINEERING AND INFORMATION SYSTEMS.

Computer, January 1986, pp. 18-30.

". . . the reader will find situations in which the term "data" is used in a context where it might be more appropriate to use the terms "information" or "knowledge".

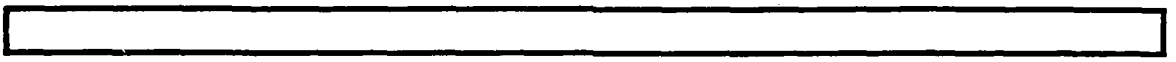
Our two principal premises are:

1) that a specific information system may involve logically and physically distributed computers and databases, and
2) that the individual functional components of an information system are driven by information from companion components and in turn provide information to other components through the transfer of data."

A gap exists between the formal disciplines in many categories of data engineering and the formal overall discipline and design methodology needed to address large-scale distributed information systems in total. Applications will not, and have not, waited for that formal discipline. Many of today's large systems were created by a partial merger of individually developed applications, or if designed more as a whole, in the absence of adequate formal, integrated design methods. Engineers must deal with the evolution of these existing systems as well as the design of new systems. Moreover, they must do so in the absence of mature design methodology.

The challenges are . . . to provide the necessary data services in the context of today's information systems. It is not feasible to redo overnight the intricate information systems essential to the operation of many segments of society. Rather, more adequate data services must be introduced and evolved without disturbing the balance of the application environment.

Architectures should proved adequate flexibility to support the growing trend to distributed systems. Information systems grow and new ones are continuously added. The architecture must support such growth.



B- Partial Solutions.

In this section, there are references to treatments of various sub-problems involved in integrating distributed, heterogeneous systems. There are technical and organizational topics which are arranged in the following sub-sections:

(1) Technical Problems: Database Management Systems.

(2) Technical Problems: System Components.

(3) Technical Problems: Methodologies.

(4) Technical Problems: Conceptual Schemata.

(5) Technical Problems: Knowledge Management.

(6) Organizational Problems: Information Systems.

(7) Organizational Problems: Organizations.

(8) Organizational Problems: Standards.

(9) Organizational Problems: Interorganizational Systems.

1) Technical Problems: Database Management Systems.

This sub-section considers problems with the conceptual and physical design of distributed or heterogeneous DBMSs, which are common concerns in the integration of heterogeneous systems.

Brodie, Michael L.

DATABASE MANAGEMENT: A SURVEY.

In On Knowledge Base Management Systems. Integrating AI and DB Technologies, Brodie and Mylopoulos (eds.), 1986, pp. 202-218.

Abstract: This chapter presents and illustrates the basic concepts of database management, outlines the major research topics, and presents the key results and future directions of database management research. Database concepts and motivations are compared and contrasted with those in Artificial Intelligence.

Cardenas, Alfonso F.

HETEROGENEOUS DISTRIBUTED DATA BASE MANAGEMENT. THE HD-DBMS.

UCLA and Computomata Intnl. Corp., October 10, 1985 (Rev. 10/23/86), 42 pgs.

"In a nutshell, the ideal long-range goals would be for a HD-DBMS to be able to support a network in which any user in any node can be given an integrated and tailored view or schema, while in reality the data may reside in one single data base or in physically separated data bases, managed individually by the same type of DBMS (by the only one the user understands) or by a different DBMS. No HD-DBMS with such full capabilities is available today. There are many unsolved problems, and others remain to be uncovered. However, major research and development projects in this arena are leading toward some partial attainment of the previous long range objectives."

This paper outlines the range of approaches to the heterogeneous challenge, from the extreme of data base unload/load, to a common interface for DBMS, to the top of the line and long range R&D and prototype efforts. Then it describes the UCLA HD-DBMS project and progress striving for the longer-range goals.

Deen, S. M.; Amin, R. R.; Ofori-Dwumfuo, G. O. and Taylor, M. C.

THE ARCHITECTURE OF A GENERALISED DISTRIBUTED DATABASE SYSTEM - *PRECI**

The Computer Journal, vol. 28, no. 3, 1985, pp.282-290.

PRECI* (pronounced presi-star) is a research prototype for a generalised distributed database management system (DDBMS). Principal architectural characteristics are:

(1) A centralised control system with heterogenous databases (including pre-existing databases) as nodes.

(2) Both inner and outer nodes to suit differing user requirements (see later).

(3) Retrieval and update facilities for global users, with full location transparency for the inner nodes.

(4) Ability to allow other new nodes to join.

(5) Ability to link with other distributed databases (external DDBs) at peer level.

(6) Maximal nodal autonomy.

Data integration, metadata and staged binding of queries may be seen as other main features.

The paper describes a five-level schema architecture, nodal and global control systems and the elements of a global query processor.

The basic design is now complete, and a pilot system is now being implemented. It is expected to include and initial version of an external protocol for linking to EEC countries over national data-communications networks. In the implementation, there will be two nodes co-existing in the same Honeywell computer, two nodes in separate VAXs, and external link to a Multidatabase project.

Elmasri, R.; Larson, J. and Navathe, S.

SCHEMA INTEGRATION ALGORITHMS FOR FEDERATED DATABASES AND LOGICAL DATABASE DESIGN.

U. of Houston, Texas; Honeywell, Minnesota and U. of Florida, Gainesville, 105 pgs.

Abstract: In recent years, schema integration has become increasingly important in light of two developments: (1) The growth of available commercial and experimental database management systems; this led to the federated database approach to dealing with the problem of numerous databases -- implemented on incompatible DBMSs -- within the same enterprise, and (2) the demand for complex databases with numerous users; logical database design methodologies were developed so that individual users could independently design their own schemas which would then be integrated into a global conceptual schema. This paper presents a methodology for schema integration that can be used in the above two contexts.

The steps of integration involve: (1) Specification of formal assertions to describe correspondences among the schemas to be integrated, (2) integration of entity sets and

relationships, and (3) generation of mapping rules that are used to transform requests between schemas. We present algorithms to collect and check the consistency of the assertions, and to perform the integration. These algorithms form the basis for a set of semi-automated tools to aid in schema integration.

Elmasri, Ramez and Navathe, Sham

OBJECT INTEGRATION IN LOGICAL DATABASE DESIGN.

Proc. of the IEEE COMPDEC Conf., April 1984, pp. 426-433.

Abstract: View integration is one of the important phases in logical database design. During this phase, the individual views designed by separate user groups within the organization are integrated into a conceptual schema for the entire organization. In this paper, we present rules for some of the aspects of view integration, namely integration of entity classes in different views. We then compare our rules with previous approaches, and discuss some of the problems which have to be solved for other aspects of view integration.

Gligor, Virgil D. and Luckenbaugh, Gary L.

INTERCONNECTING HETEROGENEOUS DATABASE MANAGEMENT SYSTEMS.

Computer, January 1984, pp. 33-43.

The article analyses existing approaches to interconnecting heterogeneous DBMS, and in the process reviews four experimental DBMS projects in which effective interconnection of such sytems was at least partially achieved.

An overall architectural model for interconnecting heterogeneous, remote DBMS is presented. The model was derived from three groups of goals that were common to various DBMS application environments: achieving transparency of local DBMS, maintaining the autonomy of all DBMS sites, and the use of standard communication protocols and existent networks. The functional layering of the model refers only to the application layer of the ISO reference model, and consists of three sublayers: (1) the global data manager (GDM) which is the top-most sublayer, and provides services directly to the end-user; (2) the distributed transaction manager (DTM), the middle sublayer, which supports the services of the GDM and requires the services of the next sublayer; and (3) the structured-data transfer protocols (SDTP), which is the lower sublayer.

The authors then present the features of the GDM among which are the global data model, query decomposition and translation, execution plan generation and results integration. The approaches of four projects (The Chemical Substances Information Network (CSIN, initiated by CCA), the UCLA DBMS project, XNDM which was initiated at the National Bureau of Standards, and Multibase by CCA) to this issue are presented.

The features of the DTM are concurrency and recovery control. SDTPs are application-level protocols required for the interconnection of remote heterogeneous DBMSs. Their main role is

the preservation of the meaning of transmitted data. The authors point to a need for a canonical format as the most appropriate method for data translation due to the ease with which a new host can be added to a network.

Goldhirsch, David and Yedwab, Laura

PROCESSING READ-ONLY QUERIES OVER VIEWS WITH GENERALIZATION.

Computer Corporation of America, Cambridge, Massachusetts, Technical Report CCA-84-03, June 1984, 8 pgs.

Abstract: The traditional Query Modification approach to query processing is inappropriate for views involving generalization. We use a combination of modification and materialization for queries over such views. Furthermore, by choosing modification or materialization as part of global optimization, we permit more optimization than would be provided by a purely modifying approach.

Heimbigner, Dennis and McLeod, Dennis

A FEDERATED ARCHITECTURE FOR INFORMATION MANAGEMENT.

ACM Transactions on Office Information Systems, vol. 3, no. 3, July 1985, pp. 253-278.

Abstract: An approach to the coordinated sharing and interchange of computerized information is described emphasizing partial, controlled sharing among autonomous databases. Office information systems provide a particularly appropriate context for this type of information sharing and exchange. A federated database architecture is described in which a collection of independent database systems are united into a loosely coupled federation in order to share and exchange information. A federation consists of components (of which there may be any number) and a single federal dictionary. The components represent individual users, applications, workstations, or other components in an office information system. The federal dictionary is a specialized component that maintains the topology of the federation and oversees the entry of new components. Each component in the federation controls its interactions with other components by means of an export schema and an import schema. The export schema specifies the information that a component will share with other components, while the import schema specifies the nonlocal information that a component wishes to manipulate. The federated architecture provides mechanisms for sharing data, for sharing transactions (via message types) for combining information from several components, and for coordinating activities among autonomous components (via negotiation). A prototype implementation of the federated database mechanism is currently operational on an experimental basis.

Keller, Arthur M.

THE ROLE OF SEMANTICS IN TRANSLATING VIEW UPDATES.

Computer, January 1986, pp. 63-73.

The author first states that shared integrated databases are often too complex for typical users to manipulate, and that views are defined for each class of user as interfaces that give them only the information that is relevant to them. He also states that handling queries expressed against views is well understood: "The users's query is composed with the view definition so as to obtain a query that can be executed on the underlying database."

Updates expressed against a view, however, have to be translated into updates that can be executed on the underlying database, the solution to which is inherently ambiguous. The author gives an overview of other proposed solutions, then presents his approach, which involves imposing syntactic criteria on the view update translations, enumerating the alternative translations that satisfy these criteria, and then, at view definition time, using semantics to choose among these alternatives.

Kent, William

LIMITATIONS OF RECORD-BASED INFORMATION MODELS.

ACM Transactions on Database Systems, vol. 4, no.1, March 1979, pp. 107-131.

"Record structures are generally efficient, familiar, and easy to use for most current data processing applications. But they are not complete in their ability to represent information, nor are they fully self-describing."

Landers, Terry and Rosenberg, Ronni L.

AN OVERVIEW OF MULTIBASE.

Computer Corporation of America, Cambridge, Massachusetts, 1982, 32 pgs.

Abstract: "This paper describes the prototype implementation of MULTIBASE -- a software system that provides a uniform retrieval interface through a single query language and database schema to data in pre-existing, heterogeneous, distributed databases. MULTIBASE efficiently

executes queries that may require data to be retrieved from different databases that have different schemata, data models, and query languages. This is accomplished without making any changes to the pre-existing databases, their database management systems, or their application programs. MULTIBASE presents the user with the illusion of a single, integrated, nondistributed database. To do this, MULTIBASE assumes complete responsibility for knowing the location of the local databases, accessing the data at each of the local database management systems, resolving data incompatiblities, and combining the data to produce a single result."

The paper is what the title says. It gives a good overview of the basic architecture, the integration of local schema, query processing and query optimization by way of examples.

Lee, Daniel T.

DECISION SUPPORT IN A DISTRIBUTED ENVIRONMENT.

National Computer Conference Proc., 1984, pp. 477-488.

A method for distributed system (DS) design with distributed databases is presented. The design satisfies geographical data independence as well as logical and physical data independence in the traditional sense.

The method is divided into two phases: The subsystem delineation phase, consisting of five steps, the result of which is a general picture of the distributed system, showing the functions of different subsystems. The distributed database development phase, in which the needs of the different subsystems, as well as possible communication between them, are defined.

Six distributed systems (IMS/MSC, CICS/ISC, Distributed INGRES, R* (R Star), Tandem's Encompass Systems and SDD-1), some which are already installed and commercially available, are examined in order to show the state of the art as well as present different issues of DS development.

The author does not fully discuss some issues such as concurrency control, failure recovery, processing requirement or software development, considering them to be beyond the scope of the paper. He also claims that the major bottlenecks in DS development are in the software development of communication systems and network database management protocols. There is no mention of heterogeneity of processors or data.

Libes, Don and Barkmeyer, Ed

IMDAS - An Overview.

Integrated Systems Group, National Bureau of Standards. (Draft)

The authors first discuss issues of data access in a computer integrated manufacturing (CIM) environment, stressing the diversity of computer systems and data organizations, the need for real-time data access, and the need for the integration of new systems into a running complex.

The approach to the problem is a unified data administration system called IMDAS (Integrated Manufacturing Data Administration System).

Characteristics of the IMDAS are a common interface to programs and a common interface to underlying commercial databases. The internal architecture resembles a hierarchy of four levels, which are distinguished primarily by scope of responsibility for data management. A description of each of the levels is presented. (See also: Barkmeyer et al.)

Litwin, Witold and Abdellatif, Abdelaziz

MULTIDATABASE INTEROPERABILITY.

Computer, December 1986, pp. 10-18.

The authors address the problem of accessing distributed heterogeneous database systems, but claim that using a global schema is too difficult an approach (and it limits the individual database autonomy), and therefore advocate what they call "interoperability". Basically, this means that the user knows he is facing multiple databases, but the system provides him with functions for manipulating data that may be in visibly distinct schemas.

The authors present a prototype multidatabase system called Multics Relational Data Store Multidatabase (MRDSM) that is representative of the interoperable approach. They discuss the system at the data definition level, showing how different databases are named, identified and accessed. Then they give examples of the data manipulation language, showing the capabilities of the system along with problems caused by semantic differences among the different databases.

Lyngbaek, Peter and McLeod, Dennis

AN APPROACH TO OBJECT SHARING IN DISTRIBUTED DATABASE SYSTEMS.

Ninth Intnl. Conf. on Very Large Data Bases, October 1983, pp. 364-375.

Abstract: This paper describes DODM, a simple model for object sharing in distributed database systems. The model provides a small set of operations for object definition, manipulation, and retrieval in a distributed environment. Relationships among objects can be established across database boundaries, objects are relocatable within the distributed environment, and mechanisms are provided for object sharing among individual databases. An object naming convention supports location transparent object references; that is, objects can be referenced by user-defined names rather than by address. The primitive operations introduced can be used as the basis for the specification and stepwise development of database models and database systems of increasing complexity. An example is provided to illustrate the use of DODM in the design of a distributed database system supporting a semantically expressive database model.

Mitchell, Mary J. and Barkmeyer, Edward J.

DATA DISTRIBUTION IN THE NBS AUTOMATED MANUFACTURING RESEARCH FACILITY.

Center for Manufacturing Engineering, NBS.

Data flows between control systems in a number of ways. Command and status information flow down and up the logical hierarchy between supervisors and subordinates, while information like workpiece location and geometry is used and modified . . . all over the facility as the workpiece chages hands.

For static data the concurrency control problem is version control. For dynamic data, there exists only one logical copy of each occurrence, and every reader must retrieve the current value before taking any action based on this data. The concurrency problem which arises in this case is the asynchronous control problem. The "active" data classes face the concurrency problems addressed by distributed database research.

The second problem is that the data dictionary functions are almost entirely human-administered at this time, and therefore tests and modifications of control programs inevitably require the foreknowledge and cooperation of the database systems programmers. . . This method of data administration will not be sufficient to perform real distributed management of the data. The development task of the next level AMRF (Automated Manufacturing Research Facility) is to automate the data dictionary functions, in other words, to manage the "metadata" - the information about the data. The primary functions to be automated are:

1) access to information on structure, location, physical mapping and DMS selection for AMRF data elements;
2) definition of user views and associated transforms by node, control process, and procedure;
3) maintenance of dynamic relationships between control programs and databases.

Navathe, Shamkant; Elmasri, Ramez and Larson, James

INTEGRATING USER VIEWS IN DATABASE DESIGN.

Computer, January 1986, pp. 50-62.

"User view integration is applicable to initial design, while existing schema integration applies to existing databases. Here, view integration refers to the activity of designing a global structure (integrated schema) starting from individual component structures (views).

We feel that view integration can be accomplished only with interactive design tools and a continuous dialog with the designer. Integration is thus somewhat subjective, with the designer helping to resolve semantic conflicts."

The authors have adopted a model called the Entity-Category-Relationship model which is and extension of the popular Entity-Relationship model. They present a view integration framework which can be interpreted both in a manual and semiautomated context. Then they discuss the view integration process as preintegration, integration of similar and different object classes, and integration of relationships.

Navathe, S. B.; Sashidhar, T. and Elmasri, R.

RELATIONSHIP MERGING IN SCHEMA INTEGRATION.

Proc. of the 10th Intnl. Conf. on Very Large Data Bases, August 1984, pp. 78-90.

Abstract: Merging of relationships among data is an important activity in schema integration. The latter can arise as integration of user views in logical database design or as the creation of a global schema from existing databases in a distributed or centralized environment. During the "view integration" phase of design, separate views of data held by different user groups are integrated into a single conceptual schema for the entire organization. In this paper we use a variant of the entity relationship model to represent schemas or user views and discuss the problem of integrating relationships from different schemas. Using three major criteria for comparing relationships, we develop a hierarchical comparison scheme. Each case represented by the terminal nodes of this hierarchy is discussed separately and rules of integration are developed. The problem is dealt with in a general sense so that the qualitative discussion is applicable to several other semantic data models. After a paper on object class integration at COMPDEC 84, this work constitutes our next step in the research on schema integration.

Ruschitzka, Manfred; Choi, Andrew and Clevenger, John L.

SIBYL: A RELATIONAL DATABASE SYSTEM WITH REMOTE-ACCESS CAPABILITIES.

National Computer Conference Proc., 1984, pp. 537-545.

Sybil is a self-contained, microprocessor-based relational database system which runs on the IBM pc using DOS enhanced by an RS232 package. It also has the ability to transfer data to and from a remote database.

The system has six components. Five form layers over DOS and are used for local DB operations: Storage manager, relation manager, query processor, query parser and command interpreter. The sixth component is a transformer which gives the remote operations capability.

The transformer is invoked for terminal emulation and, using the communications package, makes Sybil appear to be a terminal to the remote system. DB commands can then be used to

retrieve data from the remote DB. The data is stored in DOS files then, once the retrieval is complete, recalled to be transformed into the format compatible to Sybil.

The system can now be used to communicate with remote systems running INGRES and FRAMIS.

Shipman, David W.

THE FUNCTIONAL DATA MODEL AND THE DATA LANGUAGE *DAPLEX*.

ACM Transactions on Database Systems, vol. 6, no. 1, March 1981, pp. 140-173.

DAPLEX is a database language which incorporates:

(1) a formulation of data in terms of entities;

(2) a functional representation for both actual and virtual data relationships;

(3) a rich collection of language constructs for expressing entity selection criteria;

(4) a notion of subtype/supertype relationships among entity types.

This paper presents and motivates the DAPLEX language and the underlying data model on which it is based.

Su, Stanley Y. W.

MODELING INTEGRATED MANUFACTURING DATA WITH *SAM**.

Computer, January 1986, pp. 34-49.

Because the nature of a CIM (computer integrated manufacturing) system is heterogeneous, a number of database requirements must be considered:

First, data to control and support design, manufacturing, sale and service of products will be physically stored at, and processed by, the component systems. Data sharing among these component systems requires a common data model that explicitly defines the structures, constraints and operations (e.g., expert rules) that represent the semantic properties of the data.

Second, the integration of diverse activities in a factory requires that the data model be rich in semantics and capable of defining diverse engineering, statistical and business databases.

Third, the data model should provide strong data typing to allow the definition and processing of complex data types.

Existing data models -- such as relational, network and hierarchical models -- and the commercially available DBMSs based on these models are not entirely suitable for managing

databases in a CIM environment. The reason is that these data models are designed mainly for managing business-oriented rather than CAD/CAM databases.

. . . the University of Florida have investigated the requirements of managing scientific and statistical databases and have developed a semantic association model called SAM*.

This article identifies some important database requirements and semantic properties of data used in integrated manufacturing and shows how SAM* can be used to model and capture these properties. Included are problems and treatments of:

complex data types
temporal, positional and procedural relationships
hierarchicies of data structures
recursive definition of data objects
complex objects
modeling of partitioned and replicated data
naming and describing of a class of objects
modeling of multiple versions of a design
distinction between summary and category attributes

Thompson, Paul

NATURAL LANGUAGE ANALYSIS, INFORMATION MODELING, AND DATABASE ENGINEERING.

Proc. of the 18th Annual Hawaii Intnl. Conf. on System Sciences, 1985, pp. 500-514.

An overview of some concepts in knowledge engineering is presented. First is the issue of analyzing the content of user expert natural language. Five "fundamental facts" are presented:

1) The purpose of any information system is to store facts.
2) The users of an information system establish a dialogue with each other through the stored facts.
3) All communication between a user and an information system can be regarded as a simplified form of natural language sentences.
4) In order for unambiguous communication to take place, both the sender and the receiver of the message must share the same mental model.
5) 100% of the user's information problem can and must be predefined in a written formal Information Model.

Then a method is discussed for building an Information Model semantic network by connecting together the elementary pieces of the information discovered during analysis. Finally, transforming the Information Model into a database design is demonstrated.

Wiederhold, G. and Qian, X.

MODELING ASYNCHRONY IN DISTRIBUTED DATABASES.

Computer Science Dept., Stanford U., 12 pgs.

A structural database model (Wiederhold, 1980), which can be seen as a formalization of the Entity-Relationship Model (Chen, 1976), is briefly described. The primitives of the model are relations and connections.

The authors propose a new relationship type which permits describing the issues raised when information is distributed over autonomous workstations. Fully synchronous update of replicated data is not feasible in such environments. The proposed identity connection identifies data which should eventually match, but which can be in differing states for some time intervals.

2) Technical Problems: System Components.

A large, distributed, heterogeneous system is composed of hardware, software and data elements which may not be easy to integrate. This sub-section serves to illustrate the possible problems that may be faced in such an attempt.

Boari, M.; Crespi-Reghizzi, S.; Daprá, A.; Maderna, F. and Natali, A.

MULTIPLE-MICROPROCESSOR PROGRAMMING TECHNIQUES: MML, A NEW SET OF TOOLS.

Computer, January 1984, pp. 47-59.

The Multi-Micro Programming Line (MML) is an integrated tool set or development system for programming distributed applications. It is based on the host-target approach (target hardware prototype design or selection, and software development and testing on a host computer), but the target can be any multi-micro configuration out of an open set of supported architectures (i.e. heterogeneity).

The MML system has three requirements: First, that MML can be retargeted to prototype systems of any reasonable architecture (at a fraction of the initial cost). Within one architecture, configuration may vary considerably. Secondly, that prototype configuration and allocation are design variables that should be exploited to meet such requirements as real-time response or system availablility and to optimize cost-effectiveness of resources. Thirdly, that MML provide specific tools for hardware description and resource allocation.

The authors survey other current proposals and available software tools for multiprocessors. They are classified into three generations based on their support for high-level sequential programming, parallel logical processing (concurrency), and distributed computation. The first generation proposals provide support for the three concepts with separate tools. The second generation systems offer concurrent languages instead of sequential tasks coordinated by multitask kernels, but they do not provide facilities for assigning computation to processors and for handling interprocessor communication. These facilities plus a concurrent language create the third generation. All three generations support I/O operations in different ways. MML is classified as a third-generation system.

The MML system is then presented. Mentioned are sequential and concurrent features of the language, program development tools, and run-time debugging and execution control support. The first version of the system, completed at the end of 1981, has a host environment of a PDP-11/34 under RSX. All tools are coded in OMSI Pascal. The implementation produces executable programs for the Z80-based Mimp architecture prototypes. The same version also operates for Zeta boards, an industrial set based on Intel 8085 processors.

Parr, F. N. and Strom, R. E.

NIL: A HIGH-LEVEL LANGUAGE FOR DISTRIBUTED SYSTEMS PROGRAMMING.

IBM Systems Journal, vol. 22, nos. 1&2, 1983, pp. 111-127.

Network Implementation Language (NIL) is being used for the implementation of prototype communication systems. It is designed for writing executable architecture which can be compiled into efficient code for the different machines and run-time environments of a family of communicating products. Features include (1) high-level primitive type families supporting constructs needed for concurrent systems, (2) facilities for decomposition of a system into modules which can be dynamically installed and interconnected, (3) compile-time typestate checking in order to enhance language security without incurring large execution-time overhead.

The authors state that NIL is a programming language that is (1) at a suitable level of abstraction for communications architecture specification, (2) sufficiently general purpose to be suitable for programming product-specific functions not defined as part of the architecture, (3) effective for defining the configuration functions by which processes are initiated and terminated, (4) secure enough to support "open" layered systems, and (5) compilable into efficient code for complete software systems.

The features that support open design are: A security mechanism that allows user-written options to coexist with system-supplied options, strong typing of interfaces, a dynamic binding comcept and access control functions.

Shatz, Sol M.

COMMUNICATION MECHANISMS FOR PROGRAMMING DISTRIBUTED SYSTEMS.

Computer, June 1984, pp. 21-28.

The article covers applications-level software-communication mechanisms that have appeared in other literature. It focuses on communication mechanisms explicitly designed for distributed programming, addressing particularly what the various communication mechanisms do, as opposed to how they are implemented. Specific programming examples show how communication mechanisms influence the programming process.

A distributed processing system is defined as a collection of distributed processing components physically and logically interconnected, possessing a decentralized, system-wide resource control, and executing application programs in a coordinated manner. Coordination is established by communication and synchronization.

The author then discusses two major classes of communication, synchronous and asynchronous. He shows the steps involved in communication between single or multiple senders and receivers in either of the classes. He also mentions possible combinations of the classes to form selective or nonselective synchronous or asynchronous receivers.

Finally, examples are presented of several programming languages proposed for developing software for distributed-computing systems. The use of specific language-based communication primitives is demonstrated by presenting PLITS, Ada and CSP solutions (from other sources) to the producer/consumer, bounded-buffer problem.

Wah, Benjamin W.

FILE PLACEMENT ON DISTRIBUTED COMPUTER SYSTEMS.

Computer, January 1984, pp. 23-32.

The author examines the issues of initial file placement and subsequent file migration on a distributed computer system. In general, a piece of information may be processed repeatedly within a local subsystem before it changes locality. The information is usually partitioned into fragments (files). The partitioning may be vertical (attributes divided into possibly overlapping fragments) or horizontal (dividing the instances of an object into fragments).

The article discusses simple and general file allocation (with network design and query processing), file migration and file migration detection. Issues such as availability and delay constraints are also examined.

3) Technical Problems: Methodologies.

The design or development of information systems involves many steps, from the definition of the system requirements to the testing and evaluation. This sub-section presents methodologies that can be used in this process, or part of this process. Some methodologies concentrate on obtaining system design requirements, while others focus on the subsequent design, implementation and testing.

Ahituv, Niv; Hadass, Michael and Neumann, Seev

A FLEXIBLE APPROACH TO INFORMATION SYSTEM DEVELOPMENT.

MIS Quarterly, June 1984, pp. 69-78.

The Information Systems Development Life Cycle (ISDLC) is usually treated as a rigid sequence of activities. The paper presents two classes of factors that, when included in a practical procedure for planning and reviewing an ISDLC, introduce a measure of flexibility that better reflects the nature of development projects

The two classes of factors are: Project/environment factors derived from requirements imposed on the system by users and other related bodies, and factors derived from the nature of the project itself (whether it is the development of a new system or the modification of an old one, or whether it is and in-house effort rather than the purchase of a 'canned' package).

A procedure illustrated by a case study is presented, showing the effect of the factors on the various dimensions of an ISDLC.

Appleton, Daniel S.

BUSINESS RULES: THE MISSING LINK.

Datamation, October 15, 1984, pp. 145-150.

. . . the problem for data management is not managing facts. It is managing meanings. This is because for each meaning, we can have many facts. The meanings (not the facts) define a organization's (or a person's) concept of reality.

These meanings are interdependent; they are a fabric. These fabrics of meanings are not cosmically defined. They are defined by convention.

Business rules exist, whether explicitly documented or not. They are not invented by computer people. They are, in effect, statements about the basic structure of data in a business. Generally, they are described in terms of:

1) the things in a business that people need to know information about;
2) the descriptors of those things; and
3) the relationships among those things.

The things themselves are generally called entities. The descriptors are generally called attributes of entities. And the relationships among those entities (and their attributes) are generally called relationships.

A procedure for discovering Business Rules follows:

1) Identify a set of entities;
2) Identify a single attribute of each entity that allows for the identification of a unique instance of an entity;
3) Build an information model to validate and document the relationships that exist between entities;
4) Identify additional attributes of interest about each entity.

If a business's activities need to be changed, then the changes must be evaluated with respect to how the existing set of Business Rules will be affected. . . Business Rules can manage the development and evolution of shared data. . . They provide the link between high-level data architectures and project-level physical database designs; they provide the link beween data structuring and data processing; and they provide the links among shared data.

Batiste, John L. and Jung, John T.

REQUIREMENTS, NEEDS, AND PRIORITIES: A STRUCTURED APPROACH FOR DETERMINING MIS PROJECT DEFINITION.

MIS Quarterly, December 1984, pp. 215-227.

The paper presents RNP as a method for high level MIS project definition. The authors claim that it combines the more effective concepts of three other methods (CSF, BSP and ATT) while avoiding their pitfalls. Thus it does not require undue manpower expenditures while capturing the important aspects of a businesses problem.

RNP is presented in three phases: The executive session, a task force analysis and an executive presentation. The paper includes a description of the roles of the participants as well as observations as to the use of some tools and the pitfalls that the method avoids.

Hackathorn, Richard D. and Karimi, Jahangir

COMPARATIVE EVALUATION OF INFORMATION ENGINEERING METHODS.

Info. Sys. Res. Grp, College of Bus. and Adm., U. of Col. at Denver, February 1986, 39 pages.

The paper evaluates the current methodologies and tools of information engineering. (IE) Twenty-seven widely-cited methods are reviewed.

The methods are evaluated in two dimensions: Breadth, which is an extension of traditional frameworks for the system life cycle in which the overall mission and nature of the organization are included. The depth dimension deals with the extent to which an IE method is conceptual and abstract versus practical and concrete.

The authors presented five points as the conclusions of the study:

(1) No single existing tool, technique or methodology can support the entire breadth and depth of the framework as it is defined in the paper.

(2) There is a lack of "Pragmatic Planner" tools and "Conceptual Developer" methodologies.

(3) A continuous planning process should be established that aligns the development and operation of information systems to the strategic plans and directions of the organization, and exploits the advantages of the existing information technology to change the nature of competition and/or to move into new areas of business.

(4) The primary aim of most IE methods is to improve limited (i.e. to applications), global and stratgeic productivity for an organization. The last two elude most organizations because of the lack of IE methods that provide integrated coverage across the align and exploit process, as well as inadequate attention to strategic planning by top management and to information management as part of that planning.

(5) Responsibility for IE is shifting to end-users, which is degrading the stability and integrity of the overall information management of the organization. IE methods are not currently providing an integration of end-user developed systems with "mainstream" systems.

Ramey, T. L.; Brown, R. R.; Melkanoff, M. A. and Rodriguez-Oritz, G.

ELKA INFORMATION MODELING.

Hughes Aircraft Co. and UCLA, 55 pgs.

Abstract: This paper describes the Entity-Link-Key-Attribute (ELKA) information modeling approach, and its associated symbolic language, which form a technique for describing information. The ELKA modeling technique has been developed at the Hughes Aircraft Company as a part of a general methodology for constructing models of integrated engineering and manufacturing systems. The modeling technique is first presented informally in terms of its descriptors; the graphical symbolism utilized to construct ELKA information models is then introduced through an example; and a metamodel describing the basis of the ELKA modeling approach in terms of its own descriptors is presented. Finally, there is a comparison of the features of the ELKA modeling technique with those of other information and data modeling approaches that have recently appeared in the literature.

Roman, Gruia-Catalin; Stucki, Mishell J.; Ball, William E. and Gillett, Will D.

A TOTAL SYSTEM DESIGN FRAMEWORK.

Computer, May 1984, pp.15-26.

In developing integrated computer system design methodologies, several issues must be considered: The application drives the methodology. Methodologies must exploit available technology to enhance both design productivity and quality. If corporate resources (such as available manpower, training and experience of personnel or project management practices) are not taken into account during design, the resulting methodology will be severely limited.

The TSD framework is partitioned into six stages (each of which has more than one phase): problem definition, system design, software design, machine design, firmware design and circuit design. All phases are completed in ten steps: formalism selection, formalism validation, exploration, elaboration, consistency checking, verification, evaluation, inference, invocation and integration.

Comparisons are made to the system design phases proposed by other authors, and to the steps covered by other methodologies. Compatibility with DoD specification standards is shown. The authors also discuss hardware/software trade-offs in allocating system functions.

An example of developing a system design methodology for a particular company and application area outlines the strategy of going from framework to methodology. The steps are:

- Context identification.
- Framework pruning: removal of unnecessary stages and phases.
- Selection and validation of the specification language.
- Selection of design/analysis techniques.
- Sequencing of design/analysis activities.
- Addition of project management components: activities that support system design.

The authors envision the TSD framework as a means by which existing methodologies may be rigorously evaluated against each other before empirical experiments are set up. It is also the basis for a systematic approach to the development and evaluation of design methodologies tuned to the needs of particular application areas.

Shank, Michael E.; Boynton, Andrew C. and Zmud, Robert W.

CRITICAL SUCCESS FACTOR ANALYSIS AS A METHODOLOGY FOR MIS PLANNING.

MIS Quarterly, June 1985, pp. 121-129.

The article addresses the use and benefits of the CSF method in light of a CSF study conducted at Financial Institutions Assurance Corporation.

The impact of the study on the corporation was in four distinct areas: A new information system plan (information processing, office automation and future automatomation), improved corporate attitude toward data processing, increases in staff productivity and the adoption of the CSF method as a continuing management tool.

The authors explain the project's success and present guidelines for use of the CSF method in other organizations.

To be noted is the fact that the initial purpose of the study was to simply evaluate the firm's existing data processing system in light of intermediate-term corporate objectives. But the outcome was a fundamental rethinking of the nature of the corporation.

The authors also mention the conceptual nature of the CSF method, which agrees with Hackathorn and Karimi (February, 1986).

4) Technical Problems: Conceptual Schemata.

In this sub-section, we indicate some of the problems in providing and using a conceptual schema for information systems and the underlying data. The importance of the schema becomes more apparent in the next sub-section on knowledge management.

Cardenas, Alfonso F. and Wang, George R.

TRANSLATION OF SQL/DS DATA ACCESS/UPDATE INTO ENTITY-RELATIONSHIP DATA ACCESS/UPDATE.

IEEE Proc. of the 4th Intnl. Conf. on Entity-Relationship Approach, October 28-30, 1985, pp. 256-267.

Abstract: A method or algorithm is presented for translating SQL data accessing/update into ER data accessing/update. We first present the ER data manipulation language adapted to fulfill the needs of a global conceptual language for the UCLA heterogeneous distributed DBMS project briefly outlined. Different relations in a relational schema may correspond to different structures in the ER schema (entity sets and relationship sets). Thus, various translation rules are developed to handle arbitrary SQL commands on the different types of relations. The need for this translation algorithm also arises, of course, with the need to provide a relational interface to a native ER data base system.

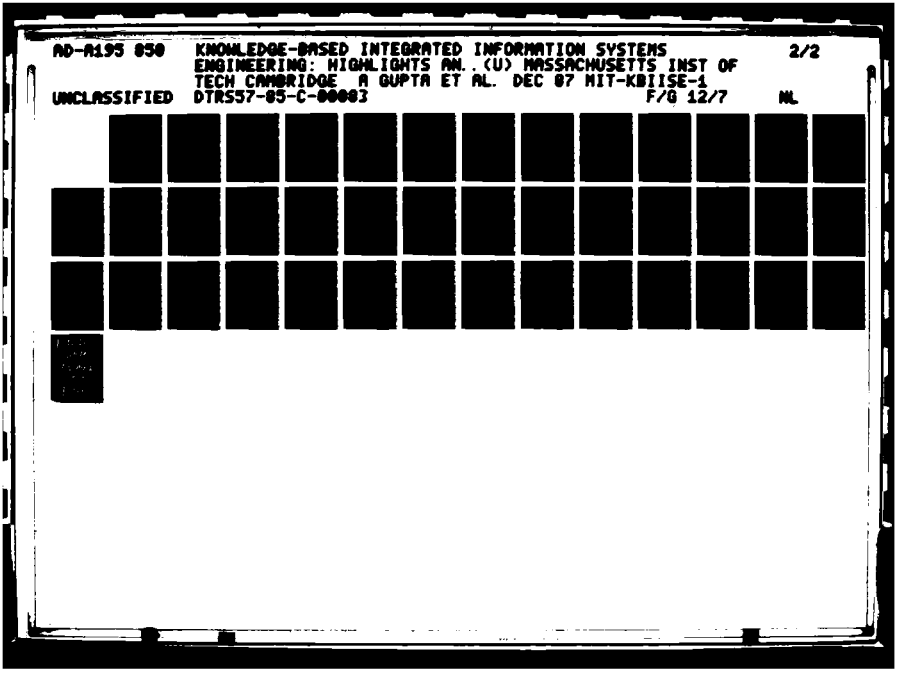
De Troyer, O.

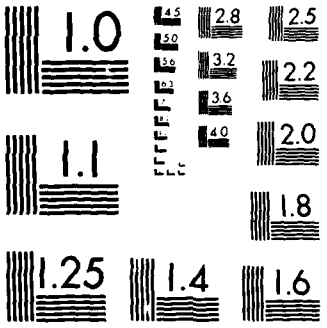
ON RULE-BASED GENERATION OF CONCEPTUAL DATABASE UPDATES.

Data and Knowledge, Proc. of IFIP TC2 Working Conf., November 1986, 19 pgs.

Abstract: Designing an information system is often a highly complex task. This complexity implies the need for computer aids. In this paper we describe how to automatically transform a certain kind of high level, understandable update specification into executable update procedures. This transformation is driven by a general rule-base. The underlying principles are independent of the data model and the manipulation language being used; we apply general methods from the area of program specification and program transformation. This allows us to prove easily the correctness of the generated procedures with respect to the input specifications.

The results obtained so far are promising for the area of software Engineering known as automated prototyping.


AD-A195 850
KNOWLEDGE-BASED INTEGRATED INFORMATION SYSTEMS ENGINEERING: HIGHLIGHTS AN..(U) MASSACHUSETTS INST OF TECH CAMBRIDGE A GUPTA ET AL. DEC 87 MIT-KBIISE-1
2/2
UNCLASSIFIED
DTRS57-85-C-00083
F/G 12/7
NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

De Troyer, O.; Keustermans, J. and Meersman, R.

HOW HELPFUL IS AN OBJECT-ORIENTED LANGUAGE FOR AN OBJECT-ORIENTED DATABASE MODEL?

Proc. of the 1st Intnl. Workshop on Object-Oriented Databases, September 1986, 20 pgs.

Abstract: One of the more interesting features of an Object-Oriented Database Model would seem to be the ability to allow proper subtyping of concepts. Now, most object-oriented languages have an "inheritance" construct. Intuitively one should be able to match these constructs with each other in some way. To get a feel of this we take one particular OO Data Model, the Binary Relationship Model and one particular OO Language, Object Pascal and try to express conceptual database update procedures. We find the results are not very satisfying to the conceptual model designer because of a deep problem: one needs an extra level of abstraction in the OO Language. The problem is not specific to the examples chosen. We discuss how it could be handled by a specific form of dynamic typing.

De Troyer, O. and Meersman, R.

TRANSFORMING CONCEPTUAL SCHEMA SEMANTICS TO RELATIONAL DATABASE APPLICATIONS.

Fourth Scandinavian Research Seminar on Information Modeling and Data Base Management, June 1985, 32 pgs.

Abstract: The ISO report on conceptual models (1982) proposed the separation of the conceptual schema from the database schema. In this paper we show how to transform a semantic network oriented conceptual schema into a relational database and how this information can be used to translate conceptual language queries into relational queries (SQL).

In doing so, we establsh a clear and explicit link between IS theory and its application areas such as software engineering support.

Meersman, R.

REASONING ABOUT CONCEPTUAL DATABASE DESIGN -- A SEMI-TUTORIAL.

Published in extended form in Proc. of IFIP TC2 Working Conf. (Data and Knowledge), November 1986, 18 pgs.

"Mostly database systems are constrained by implementation considerations such as available DBMS, programming language dependency, etc.. Generally, the major problem is the poor capability of most modeling approaches to actually (formally) represent -- and store -- knowledge in a non-procedural fashion, i. e. to become as independent as possible from the

processing that will have to be performed on that stored knowledge. Of course, no representation into any formalism can be entirely independent of the interpretation mechanism for that formalism, by definition: it always requires some kind of "syntax" which necessarily reflects the chosen (meta-)model, and furthermore depends on an external presentation for the "pure" concepts of the universe of discourse under consideration.

"By limiting the number of meta concepts however, and keeping them as "conceptual" as is (pragmatically) possible, we can hope to arrive at a method or model in which it is possible to describe many aspects of the data modeling process itself. As a consequence, we would arrive at an "object system" in which we can reason about such difficult things as database design knowledge, ontology determination, referential integrity, constraint consistency etc.."

Nijssen, G. M.

AN EXERCISE IN CONCEPTUAL SCHEMA DESIGN.

Working Paper IFIP WG 2.6 (Databases), October 1976, 42 pgs. (Preliminary version)

Summary: The discussion which data model is the best, and more particular, whether the normalized relational model or the CODASYL network model is the best data model is becoming more irrelevant as there is an increasing acceptance of the three schemata approach, one of which is the Conceptual Schema, or Universe of Discourse Shcema. The three schemata approach is a framework for a more precise discussion. One of the most essential aspects of the three schemata aproach is the set of concepts to be used in the conceptual schema. There may be more than one set of concepts which satisfies the appropriate requirements. The search for the best overall DBMS framework and the most satisfactory sets of concepts for the conceptual schema is in the center of the current interest of IFIP WG 2.6 (Databases). The discussion on the best set of concepts for the conceptual schema could be facilitated by the use of one or more non-trivial examples. The aim of this paper is to describe a few such examples.

Nijssen, G. M.

THE NEXT FIVE YEARS IN DATA BASE TECHNOLOGY.

Infotech 1977, London, 44 pgs.

Overview: Data base technology is there to support information systems. Data base managment is in principle just like COBOL and FORTRAN, a tool to be used in the design and implementation of information systems. If we want to know what is going to happen in the next five years in data base technology, I think we have to start with a generalized framework of an information system, which is understandable to the intelligent layman and independent of the current state of the art of software. With such a generalized framework of an information system, one is able to see where the long term development of the tools, such as data base management systems and application languages, is going. Another result of such a generalized framework of

an information system is a set of rules to evaluate the architecture and the various components of a data base managment system.

As the conceptual schema will play a key role in future data base technology, a set of conceptual schema concepts will be explained. Apart from its theoretical elegance, this set of conceptual schema concepts has proven superior value in practical applications. Information type and population diagrams are introduced to enhance the effectiveness of communicating on the conceptual schema.

The availability of the just mentioned generalized framework of an information system makes it possible to present a condensed, yet effective, way the main characteristics of the ANSI-1977 and CODASYL-1978 data base approach.

Nijssen, G. M.

ON EXPERIENCE WITH LARGE-SCALE TEACHING AND USE OF FACT-BASED CONCEPTUAL SCHEMAS IN INDUSTRY AND UNIVERSITY.

Proc. of the IFIP WG 2.6 Working Conf. on Data Semantics (DS-1), January 1985, pp. 189-204.

Abstract: In this paper I describe the results of using a "cookbook" approach in teaching Conceptual Schéma Design in a first year first semester University course with substantial numbers of students. I furthermore briefly describe the architectural environment of the Conceptual Schema in 4th and 5th generation Information Systems and Expert Systems as well as the concepts used in the Conceptual Schema itself.

Van Assche, F. J. M.

A SYSTEM DEVELOPMENT SYSTEM FOR DATA MANAGEMENT.

Proc. of the Intnl. Computing Symp. on Application Systems Development of ACM, March 1983, pp. 467-481.

Abstract: In this paper we highlight the basic principles for a systems development system. From the objectives of a systems development system we derive a particular pattern for automated systems development, which has been tested for more than 5 years.

Next, we apply this pattern to data management. The result is an information and data modelling system; this system is supported by software with the name IAST (Information Analysis Support Tools). It is our experience that it does often a more thorough analysis than human experts. It also performs the automated mapping from a binary conceptual schema -- expressed in structured English -- to a 5NF relational schema, while expert advice is given to tune this grouping.

Van Assche, F. and Meersman, R.

MODELLING AND MANIPULATING PRODUCTION DATA BASES IN TERMS OF SEMANTIC NETS.

Eighth Intnl. Joint Conf. on Artificial Intelligence, August 1983, 14 pgs.

Abstract: In this paper we show how semantic nets when interpreted as so-called conceptual schemata can be successfully made to serve as common denominator to two as yet disparate areas of artificial intelligence and real-life data management: Knowledge Representation and Data Base design.

First, we describe the requirments to Semantic Nets in the context of production data bases. Next we describe the automated mapping between the Semantic Net and (possibly several) DBMSs.

Further we show how this formal and machine accessible mapping allows to formulate structured English Queries, and expert rules in a data base context.

Implementation status and research directions for these ideas are given resp. suggested.

Van Assche, F.; Simons, D. and Vanhoedenaghe, M.

THE AUTOMATED MAPPING FROM A BINARY CONCEPTUAL SCHEMA TO A 5th NF DATA BASE SCHEMA.

Intnl. Ctr. for Information Analysis Services, Control Data Belgium, March 1983,

Abstract: In this paper we describe how to take a user oriented information model -- expressed in structured English -- and automatically generate a 5th Normal Form Data Base Schema with integrity constraints. To our knowledge this has never been done before.

When the information model gets larger, the importance of such an automated generation grows, since a manual generation then becomes error prone and nearly impossible.

The user oriented information model is a Binary Conceptual Schema as described in the report of the ISO TC97/SC5/WG3. The automated generation consists of the following parts: detection of inconsistencies and redundancies, analysis of referenceability, checking of the given reference types, generation of reference types and grouping of the binary relations. During the generation special attention is given to the integrity constraints and subtypes, so that a complete correspondence between the Binary Conceptual Schema and the Data Base Schema is guaranteed.

Van Griethuysen, J. J. and King, M. H. (eds.)

ASSESSMENT GUIDELINES FOR CONCEPTUAL SCHEMA LANGUAGE PROPOSALS.

ISO TC97/SC21/WG5-3, August 1985, 58 pgs.

The report lists guidelines for definition and assessment of Conceptual Schema Languages (CSL). It is aimed at:

- designers of information systems,
- developers of standards in the field of information systems, in particular those concerned with the development of reference models,
- suppliers of conceptual schema facilities.

5) Technical Problems: Knowledge Management.

This sub-section presents approaches to the topic of managing and manipulating knowledge. Some efforts consider the abstraction of data representations in DBMSs to achieve conceptual schema which are a form of knowledge. Others concentrate on the implementation issues and the tools. The main implementation issues involve the use of logic as a form of knowledge representation, logic and databases, and the coupling of expert systems and database systems.

Brodie, Michael L. and Jarke, Matthias

ON INTEGRATING LOGIC PROGRAMMING AND DATABASES.

Expert Database Systems, Proc. from the First Intnl. Workshop, 1986, pp. 191-207.

Abstract: Researchers are currently challenged to provide data management support for large-scale Knowledge Base Management Systems (KBMSs). This requires both a powerful knowledge representation scheme and efficient support for processing large amounts of complex knowledge. First-order logic frequently is proposed as a candidate for solving these problems in a uniform framework. However, logic programming and databases have fundamental differences in their respective treatment of databases. This chapter critically evaluates logic, logic programming, and, in particular, Prolog with respect to conventional database paradigms. The differences that must be resolved for the successful integration of logic programming and databases are identified.

Brodie, Michael L. and Mylopoulos, John

KNOWLEDGE BASES AND DATABASES: SEMANTIC VS COMPUTATIONAL THEORIES OF INFORMATION.

Computer Corporation of America, Cambridge, Massachusetts, Technical Report CCA-TR-85-01, August 1984, 44 pgs.

Abstract: In this paper, we take stock of AI and Database research to identify fundamental differences in goals, methodologies, and practice. These differences must be understood to achieve closer collaboration and interaction. Understanding the differences helps to identify problems to be addressed in integrating AI and Database technologies. It also suggests potential solutions.

The thesis of the paper is that AI research has been primarily concerned with the development of rich semantic theories for capturing and using the meaning of information while Database research has been primarily concerned with the formulation of powerful computational theories for efficiently storing and accessing large shared information bases. The thesis is supported by our impressionistic overviews of Knowledge Representation and Database research.

We conclude the chapter by characterizing a new technology called 'knowledge base management systems' (KBMSs) that will require a deep integration of AI and Database technologies.

Dayal, Umeshwar and Smith, John Miles

PROBE: A KNOWLEDGE-ORIENTED DATABASE MANAGEMENT SYSTEM.

In On Knowledge Base Management Systems. Integrating AI and DB Technologies, Brodie and Mylopoulos (eds.), 1986, pp. 227-257.

Abstract: Existing database management systems will be inadequate for many of the information processing applications (e.g., business and industrial automation, CAD/CAM, and military command and control) of the future. Our objective is to develop an advanced DBMS called PROBE that will handle more of the information types and intensional knowledge relevant to the new application areas. Our approach is to enhance existing DBMSs with (a) Abstract object types as the basis for defining new objects and operations and for integrating specialized processors, (b) Dimensional (space and time) concepts, which are the common characteristic of many of the new information types, and (c) Recursive predicates and queries, which provide intensional knowledge processing (deductive question-answering) capabilities. In each case, it is necessary to augment both the logical (data model, query language) components and the physical (storage structures, access methods, query processor) components of the DBMS. We describe approaches to addressing all these issues. We demonstrate via examples the dramatic performance improvement that can result if the DBMS optimizer is made cognizant of the information types and knowledge used by the application programs.

Deliyanni, Amaryllis and Kowalski, Robert A.

LOGIC AND SEMANTIC NETWORKS.

Communications of the ACM, vol. 22, no. 3, March 1979, pp. 184-192.

"An extended form of semantic network is defined, which can be regarded as a syntactic variant of the clausal form of logic. By virtue of its relationship with logic, the extended semantic network is provided with a precise semantics, inference rules, and a procedural interpretation. On the other hand, by regarding semantic networks as an abstract data structure for the representation of clauses, we provide a theorem-prover with a potentially useful indexing scheme and path-following strategy for guiding the search for a proof."

Gallaire, Hervé; Minker, Jack and Nicolas, Jean-Marie

LOGIC AND DATABASES: A DEDUCTIVE APPROACH.

ACM Computing Surveys, vol. 16, no. 2, June 1984, pp. 153-185.

"The purpose of this paper is to show that logic provides a convenient formalism for studying classical database problems. There are two main parts to the paper, devoted respectively to conventional databases and deductive databases. In the first part, we focus on query languages, integrity modeling and maintenance, query optimization, and data dependencies. The second part deals mainly with the representation and manipulation of deduced facts and incomplete information."

Hayes, Patrick J.

THE LOGIC OF FRAMES.

In Readings in Knowledge Representation, edited by Ronald J. Brachman and Hector J. Levesque, Morgan Kaufmann Publishers, 1985, pp. 287-295.

The extraordinary influence of frame representations led Pat Hayes in 1979 to take a closer look at what this movement had produced since Minsky's original work in 1975. According to Hayes -- and his arguments are convincing -- there are few positive suggestions to come out of this movement. In fact, outside of some new directions in "reflexive reasoning", there are virtually no new insights to be had at all from this line of work. One may argue with Hayes' pronouncement that "most of 'frames' is just a new syntax for first-order logic", but this was nonetheless the first serious attempt to cast work on frames in a formal light, that is, to carefully analyze in some standard logcial way the meaning of frame structures. Hayes directs his analysis mainly at KRL, but his comments apply to other frame systems and beyond to other representation schemes. In particular, Hayes comments on what it is exactly that makes something a representation language. He also points the way to subsequent work on understanding default reasoning in frame systems, as well as further work that deals more carefully with the epistemological/heuristic distinction.

Israel, David

NOTES ON INFERENCE: A SOMEWHAT SKEWED SURVEY.

In On Knowledge Base Management Systems. Integrating AI and DB Technologies, Brodie and Mylopoulos (eds.), 1986, pp. 97-109.

Abstract: It is widely agreed that any reasonable AI system has to reason. Once it is agreed why that must be so, one can go on to look at alternative conceptions of what reasoning consists of. This chapter presents an extremely partial and incomplete survey of some of the options.

Jardine, D. A.

SEMANTIC AGREEMENT AND THE COMMUNICATION OF KNOWLEDGE.

Proc. of the IFIP WG 2.6 Working Conf. on Data Semantics (DS-1), January 1985, pp. 71-81.

Introduction: With the advent of large scale interconnected computerized information systems, it is essential that the circumstances and prerequisites for effective communication of knowledge be more precisely defined. The purpose of this paper is to explore these topics.

The environment within which this study will be conducted starts with the observation of ISO/TC97/SC5/WG3 that ". . . the prime purpose of an Information System is to act as a communication vehicle among its users". That is, there is an active agent which provides information, and an active agent which receives it. Whether these active agents are humans or not changes the details of the communication and interpretation of information; however, ultimately a human is involved, whether as designer or user. Initially it will be assumed that a human user is the source of information, and another human is the receiver. The concepts used to analyse and define information and knowledge in this environment will then be extended to cases of communication where human involvement is less direct.

Kellogg, Charles

FROM DATA MANAGEMENT TO KNOWLEDGE MANAGEMENT.

Computer, January 1986, pp. 75-84.

The author first presents a brief history of logic systems, then discusses the benefits of adding a rule base and reasoning engine to a data management system. The remaining, larger part of the article is a description of a knowledge management system (KM-1). There is a description of its components as well as the steps the system goes through in responding to user questions. In particular, there are explanations of the query formulation, plan generation and query explanation steps.

Levesque, Hector J.

THE LOGIC OF INCOMPLETE KNOWLEDGE BASES.

In On Conceptual Modelling. Perspectives from AI, Databases and Programming Languages, Brodie, Mylopoulos and Schmidt (eds.), 1984, pp. 165-189.

Abstract: Some formal representation issues underlying the use of incomplete knowledge bases are discussed. An incomplete knowledge base is one that has only partial knowledge of the application domain. It is argued that a language that can refer to both the application domain and to the state of the knowledge base is required to specify and to question an incomplete knowledge base. A formal logical language with this expressive ability is presented and its semantics and proof-theory are defined. It is also shown how different the use of the language must be, depending on whether the interaction involves querying or defining the knowledge base.

Meersman, R. and Nijssen, G. M.

FROM DATABASES TO KNOWLEDGE BASES.

Infotech State of the Art Review, November 1983, 39 pgs.

"This paper discusses some paths from present-day database design and usage into the (future) area of knowledge representation. Conceivably, some ISO-type of conceptual schema might prove to be the 'missing link' from production database to expert system."

" In this paper we will describe the developments that are taking place in the database world in its move towards intelligent knowledge base systems."

The authors use an architecture of database management systems which is an extension of that published in the report of the ISO in 1982. The extension includes the aspects of metadata and derivation rules (inference rules in AI). The ISO report had extended the three-schema architecture endorsed in earlier ANSI reports to include more semantics in the conceptual schema and to have the conceptual schema completely free of computer efficiency or user interface efficiency aspects.

Minsky, Marvin

A FRAMEWORK FOR REPRESENTING KNOWLEDGE.

In Readings in Knowledge Representation, edited by Ronald J. Brachman and Hector J. Levesque, Morgan Kaufmann Publishers, 1985, pp. 245-262.

Minsky presents the idea of frames and frame systems, the basis of a generation of KR formalisms. Drawing from a number of ideas in the air at the time, Minsky sketches in broad strokes a representation architecture based on prototypes, defaults, multiple perspectives, analogies, and partial matching. In contrast to some of the other approaches represented in this collection, this results in a representational framework that seems to have much more to do with

cognitive memory models than with mathematical logic and logical inference. In fact, in an appendix to the paper, Minsky argues that the pursuit of formal logic in AI has been very misleading and that some of the demands of logic such as consistency and completeness are probably not even desirable in a representation language. On the other hand, he can only sketch somewhat impressionistically what a possible alternative would be like. This problem of vagueness and general lack of rigor has unfortunately followed many of those who pursued the frame ideas, as if the topic itself demanded a certain informal style of research.

Missikoff, M. and Wiederhold, G.

TOWARDS A UNIFIED APPROACH FOR EXPERT AND DATABASE SYSTEMS.

Expert Database Systems, Proc. from the First International Workshop, 1986, pp. 383-399.

The authors present a brief overview of the proposed approaches for the joining of expert systems (ES) and database systems (DBS). They discuss deductive databases, calling this the homogeneous approach, while the heterogeneous approach is that of joining an ES and a DBS in either a tightly coupled or a loosely coupled mode; the mode refers to the flexibility of the interaction between the ES and the DBS.

Then they turn to expert database systems (EDS) as a new approach. They present EDS criteria and discuss the structuring of knowledge, proposing knowledge clusters (klusters) as a new level of knowledge aggregation between rules and rulebases (analogous to relations as a level between tuples and databases). Klusters are intended to group rules in problem-related or context-related clusters.

Finally, there is an indication of areas in which research is needed for progress towards realizing expert database systems.

Reiter, Raymond

TOWARDS A LOGICAL RECONSTRUCTION OF RELATIONAL DATABASE THEORY.

In On Conceptual Modelling. Perspectives from AI, Databases and Programming Languages, Brodie, Mylopoulos and Schmidt (eds.), 1984, pp. 191-218.

Abstract: Insofar as database theory can be said to owe a debt to logic, the currency on loan is model theoretic in the sense that a database can be viewed as a particular kind of first order interpretation, and query evaluation is a process of truth functional evaluation of first order formulae with respect to this interpretation. It is this model theoretic paradigm which leads, for example, to many valued propositional logics for databases with null values.

In this chapter I argue that a proof theoretic view of databases is possible, and indeed much more fruitful. Specifically, I show how relational databases can be seen as special theories of first order logic, namely theories incorporating the following assumptions:

1. The domain closure assumption. The individuals occurring in the database are all and only the existing individuals.
2. The unique name assumption. Individuals with distinct names are distinct.
3. The closed world assumption. The only possible instances of a relation are those implied by the database.

It will follow that a proof theoretic paradigm for relational databases provides a correct treatment of:

1. Query evaluation for databases that have incomplete information, including null values.
2. Integrity constraints and their enforcement.
3. Conceptual modelling and the extension of the relational model to incorporate more real world semantics.

Sciore, Edward and Warren, David Scott

TOWARDS AN INTEGRATED DATABASE-PROLOG SYSTEM.

Expert Database Systems, Proc. from the First International Workshop, 1986, pp.293-305.

In expanding the power and flexibility of a database system by including an inference mechanism such Prolog, one approach is to have a Prolog system use an existing database system as a backend server. The authors argue that such a system has a poor division of labor and results in redundancy and inefficiency. They propose the construction of a single integrated system that combines both inferencing and data retrieval. The bulk of the study presents evidence for the proposal in the topics of indexing, blocking and buffer management, caching, query optimization and execution, and concurrency. Three functions identified as being needed additions to the Prolog system are disk pointers, tuple formats and sophisticated buffer management.

Van Assche, F.

EXPERT SYSTEMS, KNOWLEDGE BASES AND DATABASES.

Intnl. Center for Information Analysis Services, Control Data Belgium, February 1985, 29 pgs.

The author begins by describing expert systems and how they have been used to date. Then he discusses knowledge representation techniques and a knowledge based system architecture, stating that they belong to the common denominator between AI and DBMS technology. Next, he suggests a way in which DBMS technology can be used for expert systems by manipulating rules as if they were relations.

Then the discussion turns to enhancing DBMSs with expert system features. First, there is a presentation of the ISO 'conceptualisation' and '100 per cent' principles along with the ISO 3+1 architecture for information systems. Next, the author explains the NIAM method as an example

of the Binary Conceptual Schema approach to database design. This is compared to the AI semantic net formalism and an example is shown of how to map the conceptual schema to DBMS technology.

Finally, expectations for the future of information systems architectures are discussed.

Wiederhold, Gio

KNOWLEDGE AND DATABASE MANAGEMENT.

IEEE Software, January 1984, pp. 63-73.

The author first distinguishes data from knowledge, then gives two uses for the application of knowledge to databases: data reduction and correct query interpretation. Then he identifies and describes nine categories of knowledge (in decreasing levels of system abstraction) in which knowledge may interact with databases: enterprise-directing, focus management, application-specific, general procedural, structural, derived, data-domain, query and response control, and resiliency and recovery support.

This is followed by a discussion of methods of knowledge representation, and a brief mention as to how to choose among the different representations.

Wiederhold, Gio

KNOWLEDGE BASES.

Future Generation Computer Systems, North-Holland, 1985, pp. 223-236.

This paper defines structural knowledge (of a database), then shows how it can be used to improve query processing and response management in databases. Specific areas mentioned are query processing, summarization of data, integrity maintenance, cooperative responses and (briefly) natural language update.

In addition to structural knowledge, the author also discusses "deep knowledge", giving different alternatives for representation, as well as suggesting a frame-based knowledge structure derived from a related project (RX).

6) Organizational Problems: Information Systems.

This sub-section shows efforts at defining and solving possible organizational problems faced in the development and management of information systems.

Appleton, Daniel S.

VERY LARGE PROJECTS.

Datamation, January 15, 1986, pp. 63-70.

New information systems usually replace multiples of old systems or build on them, and are thus bigger almost by definition. Top management interest further increases the sizes of the projects and the budgets.

Very large projects defy traditional dp thinking because they cannot be accomplished by traditional functional decomposition systems development methods. Before a project is fully defined, requirements have changed, costs have increased or schedules have expanded.

The author presents a test (consisting of eleven characteristics, such as: There is a tight scope with few interfaces to existing systems, or whether a single individual has responsibility and authority for project success) to measure which projects have a better chance for successful completion. He then states that a close study of the characteristics shows that projects a e doable only if developed within an environment of three dimensions: (1) An information architecture, (2) a computer systems architecture and (3) a control architecture.

Finally, an asset-based life cycle model is presented to overcome the productivity problems of very large projects. The assets must be developed by the control architecture with the primary goal of reusing them.

Appleton, Daniel S.

INFORMATION ASSET MANAGEMENT.

Datamation, February 1, 1986, pp.68-76.

Information Resource Management (IRM) is just beginning to feel a need for asset management, driven by dramatic reductions in productivity, accompanied by even more dramatic increases in the volume and complexity of demand. More specifically, `very large projects' are appearing, signaling the need for a paradigm shift, away from the tradition of standalone systems development, and toward the asset-based life cycle model.

In addition, at the National Bureau of Standards Fifth Database Workshop (October, 1985), there was an agreement that information asset management would become the dominant theme in IRM. The workshop participants conceptualized information assets as five structures: data

acquisition, data storage, data manipulation, data retrieval and data distribution. The workshop proposed that over the next five to ten years, we would see a shift in management perspective from its current focus on what we think of as applications (bill of materials, accounts payable . . . etc.) to a future focus on the asset structures as the dominant concept of IRM.

The author defines strategic information units (SIUs) which are environments with distinct markets, control structures and sets of resources. He believes that traditional applications will eventually get bunched into such units.

There are five major differences between SIUs and traditional application projects: Users are treated as if they were customers rather than suppliers of requirements. The SIU must develop a distinct notion of what its products are, while very large projects embrace no distinct concept of product. SIUs give assets top priority in their resource commitment decisions, and their control stuctures ensure that the assets get reused with the minimum customization, while very large projects are so interested in building the application they have no time to focus on the development of assets. Very large projects tend to be subservient to support staffs while SIUs are not. SIUs are not established for a specific time interval, whereas very large projects are established with the assumption that they will cease to exist once the application has been implemented.

Appleton, Daniel S.

RULE-BASED DATA RESOURCE MANAGEMENT.

Datamation, May 1, 1986, pp. 86-99.

The experience of data administrators over the last six years in applying fundamental data administration concepts to the task of managing data assets suggests seven lessons:

(1) Data planning must define at least two levels of data structure (high-level groupings at one level, and relationships between them at another).

(2) Software implementation must be derived from two levels of data structure (logical and physical).

(3) The two levels of the planning model must be integrated with the two levels of the project (implementation) model (vertical integration).

(4) A consistent data structure among projects must also be provided (horizontal integration).

(5) Data elements in the installed base must be included in or mapped to the integration schema.

(6) The data modeling language controls both how data structures are defined and the rules for development of conceptual schemata.

(7) Rule-based data resource management must be supported by data modeling software that does a lot more than graphics (such as support for defining entities, attributes and relationships, provision for the development, integration and individual management of all four levels of data structure . . .etc.).

Barros, Oscar

EXPANDING THE SCOPE OF INFORMATION SYSTEMS DESIGN.

Proc. of the 18th Annual Hawaii Intl. Conf. on Sys. Sciences, 1985, pp. 678-686.

The author presents a conceptual model based on systems theory for an IS. Based on this, he relates the problem of determining information requirements to the alternatives for users' tasks. Then these alternatives are shown to correspond to different organization structures. Consequently he redefines the step of determining information requirements in IS development as the joint design of organization-information structures. Guidelines as to how this step could be implemented in practice are given along with a case to illustrate the new approach.

Metz, Richard

BOEING'S PC PRACTICES.

Datamation, January 15, 1986, pp. 85-88.

In response to a need for automation, as well as internal needs to share computing resources, to provide easy terminal access to data, and to add new computing systems and equipment without complicated and costly system changeovers, Boeing Computing Services (BCS) has promulgated a set of corporate information systems standards, intended to bring a focused, integrated approach to its computing activities.

Boeing Network Architecture (BNA) provides statements of direction for hardware, software and communications compatibility in and between all Boeing processing environments. The statements of direction are generated by top technical managers throughout the support divisions, and are introduced companywide by a central BNA planning council. The architecture provides a systematic method for limiting the number of vendors with whom Boeing deals. In addition, before purchasing equipment from a vendor, BCS asks for four basic commitments: a published architecture, a strategic products briefing, a current implementation of IBM's SNA or a short-term plan to do so, and a long-term plan to adopt the International Standards Organization's (ISO) Open Systems Interconnect (OSI) model.

There is a goal of building a hierarchy of logical communications protocols necessary for application-to-application communications, allowing workstations to communicate with hosts on a peer basis. To this end, Boeing has initiated the technical office protocols (TOP) project, involving several companies and vendors, and based on the ISO's OSI model.

BCS is trying to prevent "islands of functionality" where a highly functional pc generates reports or analyses that must be reentered at another location. Thus BCS has established a central program to stay in front of pc introductions as much as possible, setting policies and standards, and keeping up with the industry to maintain an active evaluation program.

The cornerstone of Boeing's personal computing program is the personal computing store concept, which provides all the services of a first-rate computer store, and integrates them into the corporate environment. There are provisions for system installation, maintenance, inventory and configuration control, as well as stocks of standard pcs and peripherals along with demonstration centers. Finally, there is the Boeing Personal Computing Catalog to make it easier for users to be self-sufficient.

Pratt, S. J.

APPLICABILITY OF DECENTRALIZED/CENTRALIZED CONTROL PROCEDURES IN DISTRIBUTED PROCESSING SYSTEM DEVELOPMENT AND OPERATION.

IEEE Trans. on Eng. Mgmt.,vol. EM-32, no. 3, August 1985, pp. 116-123.

The author presents an argument for the advantages of centralized control (distributed processing does not necessarily imply distributed control) of distributed systems in both the managerial structure (and thus the design and implementation) and the operation.

Centralization is to be applied at all levels, thus insuring the advantages from the lower levels to the system as a whole. The heads of the subsystems at any level form the centre of control at that level.

The advantages are seen as: unifying objectives of the system; avoiding possibly catastrophic failures and improving productivity through advance planning that takes all circumstances into consideration (which is only possible through centralized coordination or control); and insuring an orderly and successful recovery in the event of a failure.

7) Organizational Problems: Organizations.

This group of papers contains issues concerning organizations as entities, such as effects of information systems on corporate strategy, or the effect of organizational structure on information system development and vice versa.

Bakos, J. Yannis and Treacy, Michael E.

INFORMATION TECHNOLOGY AND CORPORATE STRATEGY: A RESEARCH PERSPECTIVE.

MIS Quarterly, June 1986, pp. 107-119.

This article demonstrates how detailed explanatory models of the impact of information systems on strategic performance can be created using existing theory. It also presents three perspectives of the strategic impact of information systems (internal, competitive and business portfolio), and shows how each identifies different issues of importance and suggests different theoretical frameworks as the basis for research.

The article contains useful references that discuss the impact of information technology (within a corporation) on corporate strategy and vice versa.

Emery, James C.

INTEGRATED INFORMATION SYSTEMS AND THEIR EFFECTS ON ORGANIZATIONAL STRUCTURE.

In Information Systems and Organizational Structure, E. Grochla and Szyperski (eds.), 1975, pp. 95-103.

"A system, by definition, is composed of a set of related tasks directed toward a common goal. A system is created as a way of dealing with complexity. A large task, such as managing a business organization, is too complex to handle in a monolithic way; consquently, it is broken down into a hierarchy of subtasks.

The extent to which the components of a system should be integrated is one of the basic issues facing the managers of the system. Integration involves closer coordination, tighter linkages, and greater resource sharing among components. This brings with it certain efficiencies; but it also entails costs, delays, and added complexity. Ideally, we should choose the degree of integration that best balances these opposing effects."

Gibson, David V.

CORPORATE CULTURE AND THE MAVERICK DIVISION: TWO ORGANIZATIONAL FACTORS THAT INFLUENCE THE USE AND INTEGRATION OF INFORMATION SYSTEMS.

Proc. of the 18th Annual Hawaii Intl. Conf. on Sys. Sciences, 1985, pp. 694-704.

The paper discusses how corporate cultures and subcultures impact the use and integration of computer based information systems. There is a focus on "civil wars" between centralized and decentralized data processing. The author also discusses "goodness of fit" between corporate cultures and information systems, maverick divisions and the entrepreneurial spirit, as well as the resulting dilemmas faced by information systems managers.

King, William R.

STRATEGIC PLANNING FOR IS: THE STATE OF PRACTICE AND RESEARCH. - editor's comment.

MIS Quarterly, June 1985, pp. vi-vii.

The author states that SPIS is currently popular and is so widely adopted so rapidly, that it is useful to attempt to identify the desirable directions for further growth. He declares three criteria for SPIS processes to be at a high level of sophistication:

(1) Explicitly incorporating processes for relating IS strategy to the existing business strategy of the enterprise in the operational sense, such that a significant change in business strategy would require a significant change in IS strategy.
(2) Explicitly incorporating processes for assessing the existing and planned IS resources of the organization with the objective of identifying potentially useful changes in business strategy, tactics, or processes that they may support.
(3) Incorporating the notion of information and IS as a strategic resource or competitive weapon, and explicitly involving processes for the identification of opportunities for the use of the information resource in this fashion.

He then urges the evaluation of the processes that are being used for SPIS in order to avoid the negative reaction that can occur when planning is overdone. The evaluations can take either of two forms:

(1) Evaluations performed within the individual firm to comprehensively identify the costs and benefits that are associated with its SPIS activity.
(2) Evaluations conducted across firms by identifying various levels of SPIS activity and relating them to profitability or some other measure of business effectiveness.

Rackoff, Nick; Wiseman, Charles and Ullrich, Walter A.

INFORMATION SYSTEMS FOR COMPETITIVE ADVANTAGE: IMPLEMENTATION OF A PLANNING PROCESS.

MIS Quarterly, December 1985, pp. 285-294.

The authors developed and implemented a five-phase planning process to identify and evaluate strategic information systems (SIS, the use of information systems as competitive weapons) and to win top management support. Underlying their approach is a conceptual framework that views an enterprise's suppliers, customers and competitors as the strategic targets of five strategic thrusts: differentiation, cost, innovation, growth and alliance. Strategic thrusts represent the fundamental link between the firm's strategy and its use of information technology. SISs support and shape the organization's strategic thrusts.

The paper presents the conceptual framework and the SIS planning process at GTE which puts it into practice.

Rockart, J. F. and Scott Morton, M. S.

IMPLICATIONS OF CHANGES IN INFORMATION TECHNOLOGY FOR CORPORATE STRATEGY.

Interfaces, vol. 14, no. 1, January-February 1984, pp. 84-95.

"Beyond using information technology to support the existing business strategy, firms have the opportunity of using information technology, proactively, to create new opportunities for the business. These new strategic opprotunities are being created by a broad range of information technologies. They not only lead to new markets and new products but also provide whole new ways to manage the firm."

The authors discuss three eras of use of computer-based systems. In the first two, companies were simply enabled to process paperwork faster and more accurately, while the third focuses on providing information to middle and top management.

Finally, a conceptual model of technology impact on organizational strategy is presented along with a case study.

8) Organizational Problems: Standards.

This section considers standards in different information system components. This is classified as an organizational problem since the process of agreeing upon and disseminating a standard is usually of the organizational type, even though there may be a technical challenge in developing the standard itself (as in Bachman and Ross, 1982).

Bachman, Charles W. and Ross, Ronald G.

TOWARD A MORE COMPLETE REFERENCE MODEL OF COMPUTER-BASED INFORMATION SYSTEMS.

Computers & Standards, vol. 1, 1982, pp. 35-48.

The paper identifies three information system architectures, including the ISO Reference Model of Open Systems Interconnection, and then proposes a new combination of them as a single architecture. The two other models are the ANSI/SPARC database management architecture and a model described by Bachman (January, 1980) as an attempt to define an integrated reference model of information systems.

Features of the extended model are: a single layer (formed by merging two layers of the component models) which supports interprocess communications, data storage and retrieval, and operations on the data local to a process; and a new layer (or sublayer within the application layer) which is concerned with those application management functions that are not application type specific (e.g. concerned with application process integrity or security). Diagrams of all the models are shown.

The authors strongly recommend the development of a reference model similar to that developed in this paper, and that the existing subcommittees within the ISO be altered to a form that would closely correspond with the structure of the reference model they adopt.

Bell, C. Gordon

STANDARDS CAN HELP US.

Computer, June 1984, Special Feature, pp. 71-79.

The author states that the lack of standards impedes technological progress and lowers productivity, and that the role of standards in today's computer evolution is critical.

He presents eight levels of integration that form strata in which a large set of product-segmented industries are organized:

* Discipline and profession-specific application (e.g. CAD)

* Generic application (e.g. word processing, electronic mail)
* Third-generation procedural programming languages
* Operating system
* Electromechanical
* Printed circuit board
* Standard chip
* Silicon wafer

Some standard dos and don'ts:

1) Either make the standard or follow the standard.
2) Be prepared to react quickly and follow when the de facto standard changes.
3) Change the standard when it's wrong.
4) Make somebody [person(s) or organization(s)] responsible for defining, implementing, and maintaining a standard.
5) Minimize the number of organizations responsible for a standard.
6) Remember that almost any standard is far more important than a highly refined optimum. To make progress, we often have to regress.
7) Provide and plan for evolution; it's often the fastest way.
8) Base the standard on experience, not on a committee design; if you haven't lived with a proposed standard, don't adopt it.
9) Make the standard precise, understandable, applicable, and useful at several levels of detail.
10) Remember that only one or a few standards are needed for the same function; a standard should aim toward unifying a set of alternatives. Ideally, a standard should define the interface between sets of parts, not just two parts.

He then discusses critical standards for the next generation in most of the eight levels of integration above, and concludes with a strong recommendation for a Compcon on standards.

Fairbairn, D.

COMPUTING STANDARDS - THE LAST FOUR YEARS.

Computers & Standards, vol. 1, 1982, pp. 9-12.

The author begins by mentioning a report by the National Computing Centre (UK), published in 1977, on standards in computing. The report concluded that:

- although a significant number of computing standards existed, they were little used
- the pace of standardisation was slow
- few workers in computing were aware of the existence of many standards
- testing of many products against standards is littled practised

The author then considers progress during the four years following the report. He states that the advent of distributed processing demands adherence to standards at various levels, yet standards have been defined, so far, by closed groups such as the airlines (booking system) or the Bankers Automated Clearing Services (magnetic tapes), not in open standardisation bodies such as the ISO. In addition, definition of standards is not enough to ensure adoption, but they should be:

-- the subject of wide publicity both during and after definition

-- produced to a defined timescale

-- assured of a body or bodies capable of absorbing the costs of first-off implementations

-- capable of being tested

-- in the care of responsible and responsive public bodies

Finally, he mentions validation procedures (i.e. testing against standards), the extra computing cost due to missing standards, and the UK initiative in this area.

Sirbu, Marvin and Hughes, Kent

STANDARDIZATION OF LOCAL AREA NETWORKS.

Dept. of Engineering and Public Policy, CMU. Presented at the 14th Annual Telecomm. Policy Res. Conf., April 1986.

The paper presents a detailed historical account of the standardization of local area networks.

Using the model developed by Sirbu and Stewart (1985), the authors argue that the failure to agree on a single standard could have been predicted. They also introduce a model which views the standardization decision as a problem in contingent contracting, which explains the divergent attitudes towards standardization of many of the participants.

Other conclusions are:

- Standardization activities are increasingly being undertaken as a means of clarifying areas of new technology which are poorly understood, both technically and from a marketing perspective. No analysis of standards which assumes that all parties have an equal comprehension of the subject matter and differ only in their economic interests can adequately explain the behavior of the participants. As standards become more frequently developed in advance of well defined market demand, the process comes to resemble the act of innovation in which firms struggle to develop new technologies to satisfy unclear needs. Firms frequently misapprehend either the technology or the market or both. The complexity of the issues being addressed means that much of the effort in

the development of standards lies in the process of educating the participants to a common perception of the problem to be resolved.

- While some standards such as the QWERTY keyboard layout or NTSC television have remained unchanged for decades, local area network "standards" are best understood as constantly evolving concepts. The balloting and publication of local area network standards documents has neither stopped the invention of new LAN technologies, nor stopped standards committees from considering an undending series of additions, variations and new options, both within and outside of the framework it has used to date.

- Firms differ in their degree of sophistication regarding the standards process itself. As with any bargaining situation, firms which invest time and money in preparation and mastering of the process can influence it in directions favorable to their interests.

Sirbu, Marvin and Stewart, Steven

MARKET STRUCTURE AND THE EMERGENCE OF STANDARDS.

Dept. of Engineering and Public Policy, CMU, April 1985 (rev. Oct. 1986).

A model for predicting the emergence of standards in a particular product market is presented. The key variable in the model is whether the purchase or production decision making is centralized or decentralized. There are four categories: When buyers are related and manufacturing is centralized, there is no pressure for standards to emerge. When buyers are related and manufacturing is decentralized, there is pressure, but it is only limited for unique standards to emerge. The highest pressure for unique standards occurs when buyers are unrelated, whether manufacturing is centralized or not.

Three hypotheses are thus presented:

H1: Standards are unlikely to emerge where a single buyer can achieve compatibility by buying products from a single vendor.
H2: A single standard is likely to emerge where the products which must interwork are bought by unrelated buyers.
H3: Multiple standards are likely to emerge where there is a single buyer, and where the products which are to interwork are not available from a single vendor.

The hypotheses are tested in the context of the modem and local area network markets.

The results of the two cases (as well as other products which are presented briefly) follow the predicted patterns. A number of cases which seem not to follow the predictions of the model are identified. The discrepancy is attributed to the fact that the model does not involve the process by which firms agree to a single standard (or which product will be the standard) nor does it involve the length of time it should take. Therefore, the authors assert that standards will predominate in those cases in the future.

The results also suggest that monopsony buyers can impose standards on manufacturers even where they would not otherwise be likely to emerge.

Sirbu, Marvin A. and Zwimpfer, Laurence E.

STANDARDS SETTING FOR COMPUTER COMMUNICATION: THE CASE OF X.25.

IEEE Communications Magazine, vol. 23, no. 3, March 1985.

The authors first briefly discuss standards and the standards setting process in computers and telecommunications. Then they present an account of the historical development of the X.25 packet switching standard. They evaluate the X.25 in terms of the extent of its use and its "technical quality", presenting the pros and cons.

Finally, they present seven lessons learned from the case of the X.25:

1. An imperfect standard is better than none.
2. Build a coalition in private meetings with other vendors and submit a joint proposal to formal institutional standards meetings.
3. Standards agreements for new technologies are likely to be most easily reached after there has been some practical experience with the technology, and when a number of implementations are imminent. Organizations whose implementations are imminent are the ones most motivated to develop standards.
4. Managers who can commit their organizations will be more successful in securing adoption of a standard than technical experts without decision-making responsibility.
5. Standards agreements are more likely to be reached for "natural" interfaces, that is, interfaces where there is a natural division in responsibility for the supply of equipment or services.
6. The layered approach speeds the development of standards.
7. Standards of higher quality and greater generality are likely to be produced if standard-setting organizations cooperate with each other during the development and drafting of new standards.

It is interesting to compare these lessons to the recommendations given by C. G. Bell (1984) and D. Fairbairn (1982).

9) Organizational Problems: Interorganizational Systems.

In this section, we consider the formation of interorganizational information systems and their effects on strategic goals and relationships between organizations.

Allen, David

NETWORK EXTERNALITIES AND CRITICAL MASS IN NEW TELECOMMUNICATIONS SERVICES.

Sixth Intnl. Conf. Forecasting and Analysis for Business Planning in the Information Age, Tokyo, Japan, November 1986, 25 pgs.

"Until the past few years the estimation of demand for telecommunications services could be based, in many industrialized societies, on the behavior of relatively mature networks. Since phenomena related to network externalities come into play during startup phases but much less so during maturity, questions about these externalities have received relatively little attention for purposes of demand forecasting. In the last few years, however, with the introduction of entirely new kinds of networks the situation has reversed. These new networks must of course survive startup and with a new focus on startup the externalities become a factor in the estimation of demand."

"This paper suggest a general analytic approach to the description of network externalities in telecommunications services, particularly for purposes of estimating demand. This approach, it turns out, is applicable to large-scale, PTT-based decisions such as the French videotex project and, at the same time, equally applicable to marketplace competitions between small or large firms in liberalized regimes, such as in the US, Japan or the UK."

Barrett, Stephanie S.

STRATEGIC ALTERNATIVES AND INTER-ORGANIZATIONAL SYSTEM IMPLEMENTATIONS: AN OVERVIEW.

Journal of Management Information Systems, vol. 3, no. 3, Winter 1986, pp. 5-16.

Abstract: The potential strategic importance of information systems is now an accepted fact. The context of this importance, however, transcends traditional data processing, user functional departments, and, more importantly, organizational boundaries. Inter-organizational information systems (IOS) that automate the flow of information interchange across organizational boundaries are made possible by technological advances in data communciations. This paper discusses a range of strategic options and IOS implementations.

Barrett, Stephanie and Konsynski, Benn

INTER-ORGANIZATION INFORMATION SHARING SYSTEMS.

MIS Quarterly, Special Issue, 1982, pp. 93-105.

Abstract: The authors discuss concepts underlying the growth of Inter-Organizational Information Sharing Systems (IS*). A classification scheme is presented and the issues of cost commitment, responsibility, and complexity of the operating environment are discussed. Organizational impacts associated with IS* participation are also examined.

Cash, Jr., James I. and Konsynski, Benn R.

IS REDRAWS COMPETITIVE BOUNDARIES.

Harvard Business Review, March-April 1985, pp. 134-142.

"Though high-speed data exchange between companies has been possible for many years, only lately have corporations begun to use this link strategically. Such electronic pathways usually join a buyer and a seller, and their chief use is in making a sale of goods and services. Despite their potentially great effects, in many companies clerks rather than managers make the decision to participate in these systems. An interorganizational system (an IOS) can give a corporation an edge over its competitors, but an IOS can also shift the balance of power between the supplier and the buyer, making for a possible unhealthy dependence on either side. The authors discuss frameworks to help managers weigh the strategic benefits and possible costs of an IOS for a company and for its competition. They also show the various ways an IOS can operate as well as the effects of various levels of participation in such a system."

Estrin, Deborah L.

INTER-ORGANIZATIONAL NETWORKS: STRINGING WIRES ACROSS ADMINISTRATIVE BOUNDARIES.

Computer Networks and ISDN Systems, vol. 9, 1985, pp. 281-295.

"The fundamental difference between computer-communication networks that operate across administrateive boundaries and more traditional interorganizational communication modes is

that a user in one organization can cause some event to occur automatically within the domain of another organization, without any human intervention or auditing.

We discuss four categories of network application, each of which provides a different range of capabilities to external users and a different degree of control over external-usage to an organization."

Estrin, Deborah

CONTROLS FOR INTERORGANIZATION NETWORKS.

IEEE Transactions on Software Engineering, vol. SE-13, no. 2, February 1987, pp. 249-261.

Abstract: Interorganization computer networks support person-to-person communication via electronic mail; exchange of cad/cam data, software modules, or documents via file transfer; input to an order-entry or accounting system via a database query and update protocol; and use of shared computational resources via an asynchronous message protocol or remote login. In most such interorganization arrangements, the set of resources that an organization wants to make accessible to outsiders is significantly smaller than the set of resources that it wants to remain strictly-internal (i.e., accessible to employees of the organization only). In addition, because the potential user is a person (or machine) outside the boundaries of the organization, the damage associated with undesired use can be high. Because of these characteristics, Interorganization Networks (ION's) have unique usage-control requirements.

This paper describes a conceptual model for implementing usage control in ION's. In the first half of the paper we review usage control requirements in networks that cross organization boundaries. Our analysis suggests that category sets and nondiscretionary control mechanisms can be used to isolate strictly-internal facilities from ION facilities, and distinct ION's from one another. The second portion of the paper focuses on the problem of authentication in ION's, which is an essential component of the proposed control mechanisms.

Weick, Karl E.

EDUCATIONAL ORGANIZATIONS AS LOOSELY COUPLED SYSTEMS.

Administrative Science Quarterly, vol. 21, March 1976, pp. 1-19.

"In contrast to prevailing images that elements in organizations are coupled through dense, tight linkages, it is proposed that elements are often tied together frequently and loosely. Using educational organizations as a case in point, it is argued that the concept of loose coupling incorporates a surprising number of disparate observations about organizations, suggests novel functions, creates stubborn problems for methodologists, and generates intriguing questions for scholars. Sample studies of loose coupling are suggested and research priorities are posed to foster cumulative work with this concept."

Wellman, Barry

NETWORK ANALYSIS: SOME BASIC PRINCIPLES.

In Sociological Theory, 1983, Chapter 6, pp. 155-200.

"Network analysis is a fundamental approach to the study of social structure. This chapter traces its development, distinguishing characteristics, and analytic principles. It emphasizes the intellectual unity of three research traditions: the anthropological concept of the social network, the sociological conception of social structure as social network, and structural explanations of political processes. Network analysts criticize the normative, categorical dyadic, and bounded-group emphases prevalent in many sociological analyses. They claim that the most direct way to study a social system is to analyze the pattern of ties linking its members. By analyzing complex hierarchical structures of asymmetric ties, they study power, stratification, and structural change."

C- Helpful Concepts and Tools.

This section presents miscellaneous efforts which are not directed at any of the aforementioned sub-problems in particular, but contain helpful related concepts of a technical or organizational nature. There are two subsections:

1) Use of Expert Systems.

2) Miscellaneous Concepts.

1) Use of Expert Systems.

Expert systems, or AI in general, definitely have a place among the solutions to the aformentioned problems. Some concepts already identified include expert database systems, or expert systems to aid in solving semantics problems in heterogeneous databases.

Fisher, Edward L.

EXPERT SYSTEMS CAN LAY GROUNDWORK FOR INTELLIGENT *CIM* DECISION MAKING.

Industrial Engineering, March 1985, pp. 78-83.

The author first defines expert systems and descirbes how they differ from conventional problem-solving techniques. He gives examples of some successful commercial systems, such as XCON, DIPMETER ADVISOR and MACSYMA. Then he discusses expert systems in CIMS, pointing out criteria for suitability of a task for expert system technology, such as narrow focus of the problem, no reliance on external sensing (until ES technology advances are made) and the availability of an expert to participate in the knowledge base formulation. He warns of problem areas that require the solution of more than one task type, categorizing tasks into planning, design, interpretation, diagnosis, monitoring and prediction (prime problem types for ES application are diagnosis and structured selection). Finally, he compares the project of building an expert system to conventional software projects, and mentions some systems in the manufacturing domain that are still in the research stage (ISIS, GARI and FADES).

Shannon, Robert E., Mayer, Richard and Adelsberger, Heimo H.

EXPERT SYSTEMS AND SIMULATION.

Simulation, June 1985, pp. 275-284.

The authors first describe the fields of artificial intelligence and expert systems. They discuss knowledge categories (global data, knowledge base and control or inference structure), giving examples of how each may be implemented.

Since simulation languages and simulation models contain many of the ideas being used in AI (such as the ability of entities to carry or dynamically change attributes), the authors then discuss the differences between an expert simulation system and a well-written simulation model. The main differences lie in the way the model is constructed and run, the ability to modify knowledge (in an AI-based system) in one of the three categories mentioned above without affecting the others, the nature of the data bases used and other characteristics, as well as the way in which the system and model are used.

Following that is a discussion of criteria for suitablility of applying AI to a problem (see Fisher 1985 for a comparison), as well as a brief discussion of fifth generation languages.

Finally, the authors present ideas for an expert simulation system for manufacturing, discussing the domain knowledge base, design of an experiment to be run with the resulting

simulation model, building the model, specifying input to the model, and how to run the experiment.

Shaw, Michael J.

A KNOWLEDGE-BASED FRAMEWORK FOR DISTRIBUTED DECISION SUPPORT SYSTEMS.

Proc. of the 16th Annual Pittsburgh Conf. on Modeling and Simulation, vol. 16, part 3, April 1985, pp. 831-836.

A knowledge-based framework for distributed DSS is presented. Each DSS node consists of four components: a knowledge base storing knowledge about the problem domain, a language system interacting with the users, a problem processing system and a network communication system. The knowledge of a node is represented at three levels: data (declarative knowledge about the domain, serving as the state description), knowledge-base (procedural knowledge) and control (control and problem-solving strategies). The decision support process is then treated as state-space problem solving and the goal is to have the problem solved. When the problem-solving requirements are beyond the capability of a node, the DSS invokes a distributed problem-solving procedure.

Smith, John Miles

EXPERT DATABASE SYSTEMS: A DATABASE PERSPECTIVE.

Expert Database Systems, Proc. from the First International Workshop, 1986, pp.3-15.

An expert database system (EDS) involves a combination of database management system (DBMS) and expert system (ES) technology. It is defined as "a system for developing applications requiring knowledge-directed processing of shared information". The ES is used to perform intelligent processing of information being stored in, or retrieved from, the DBMS.

The paper reviews the fundamental objectives of DBMS and ES systems. Search is the common underlying function of the two types. The inappropriate combination of two powerful search engines, however, can lead to a multiplicative explosion in computation time. It is argued that performance is the key challenge in EDS development.

A simple EDS architecture is proposed. Three kinds of system components are between the users and the information: ES's for knowledge-directed processing, a DBMS for shared information management, and specialized processors for handling special format data. The key architectural question is identified as the distribution of system functionality between the DBMS and ES components. It is proposed that the distribution criteria should be based on maximizing overall system performance.

The two principal performance bottlenecks are identified as inferential search in ES's and access to secondary storage devices in DBMS's. The two areas are explored, and the design criterion stated is that the ES should only be used for those cases of inferencing where the power

of its search mechanism is really needed. In other cases, simpler search mechanisms should be used. In particular, the DBMS should be delegated maximum responsibility for searching shared information. The goals are to minimize the amount of knowledge involved in the search cycle of the ES, and to give the DBMS maximum scope for optimization, so as to reduce secondary storage access costs.

2) Miscellaneous Concepts.

The design and implementation of a distributed, heterogeneous system will bring forth a multitude of possibly unforseen problems. Various tools were seen to be helpful in reaching solutions to those problems, such as database machines or the reconciliation of database views and object-oriented programming.

Barabino, G. P.: Barabino, G. S.; Bisio, G. and Marchesi, M.

AN INFORMATION ANALYSIS APPROACH TO THE DESIGN OF DATABASES FOR VLSI CAD-SYSTEMS.

Presented at the 4th Jerusalem Conference on Information Technology.

Abstract: The generation of a data base structure to be used by a set of CAD tools for VLSI design has been undertaken using an Information Analysis approach.

From a natural language description, an Information Structure Diagram (ISD) is generated, according to "Nijssen Information Analysis Method" (NIAM). Using software tools (CDC-IAST) the ISD is then checked for correctness and translated into a data base conceptual schema, containing all the information organized in 4th normal form, with the indication of all the constraints to be fulfilled to maintain data integrity.

Two different activities concerning the design of VLSI systems, at layout and logical gate level, have been considered. A hierarchical approach has been followed in both design activities. The conceptual data base schema for layout design has been implemented using a relational DBMS (Ingres).

Chen, Bo-Shoe and Yeh, Raymond T.

FORMAL SPECIFICATION AND VERIFICATION OF DISTRIBUTED SYSTEMS.

IEEE Trans. on Software Eng., vol. SE-9, no. 6, Nov. 1983, pp. 710-722.

The paper presents an event-based model to specify formally the behavior (external view) and the structure (internal view) of distributed systems. An Event-Based Specification (EBS) language is presented, with behavior specifications, structure specifications and verification techniques illustrated through examples. The approach is compared to Temporal Logic and Trace approaches and an extended BNF definition of the syntax of the language is provided.

Safety and liveness properties of distributed systems are specified in the model using "precedes" (time-order) and "enables" (causality) relationships among events.

The advantages of the EBS language are stated as: formality, generality, accuracy and orthogonality (i.e. properties are specified separately).

Colter, Mel A.

A COMPARATIVE EXAMINATION OF SYSTEMS ANALYSIS TECHNIQUES.

MIS Quarterly, March 1984, pp. 51-66.

The paper presents a set of dimensions or factors for the comparative examination of systems analysis methods.

Two approaches to systems analysis methods are first identified: The traditional (Couger, 1973), which concentrates on improving coding practices; and the structured (Weinberg, 1978), which also includes system design and system component interrelationships. Then the factors for the examination are presented. They include issues representing both traditional and structured analysis approaches as well as general dimensions of flexibility in level and perspective. This is followed by a review of the evolution of analysis techniques along with an evaluation of a set of techniques. Finally, recommendations are given for choosing an analysis method.

Kambayashi, Yahiko

A DATABASE MACHINE BASED ON THE DATA DISTRIBUTION APPROACH.

National Computer Conference Proc., 1984, pp. 613-625.

In typical computer systems, data are transmitted to components or VLSI circuits that perform specific database operations. This is the function distribution approach. The problems are that data transmission cost is very high and some circuits may become very slow when the data size exceeds the maximum size handled by the circuits. There are two possible solutions: operations can be performed during transmission so that the effective transmission time is reduced; or a suitable approach can be found at less communication cost. The data distribution approach is an example of the latter solution.

In the data distribution approach, each component circuit can store data and realizes most required operations. In addition, all data are divided into each component circuit so that there is a simple procedure for realizing an operation on all the data by performing corresponding operations at each component circuit separately. The approach is suitable for database operations since they have the following properties:

1. Each operation is rather simple.
2. Each operation can be realized by operations applied to subsets of data.
3. The number of data to be handled by each operation varies from very small to very large.
4. Even if the volume of data is very large, usually, a query or a modification operation needs only a portion of the data, selected by some specified criteria.

5. There are operations like join, sort, and search, which are time-consuming.

Component circuits in which subsets of the tuples of a relation can be stored are called functional storage circuits. These circuits along with index circuits compose a hardware file system, which has advantages over conventional file systems: frequent rebalancing is not required, parallel processing is possible, and various operations can be realized.

Finally, the author presents an organization of the VLSI functional storage circuits that maximizes performance. The organization is not the best by current technology, but the author believes it can be realized in the near future.

Ma, Richard Perng-Yi

A MODEL TO SOLVE TIMING-CRITICAL APPLICATION PROBLEMS IN DISTRIBUTED COMPUTER SYSTEMS.

Computer, January 1984, pp. 62-68.

Efficient utilization of distributed computer systems demands a cost-effective interconnection network, an application requirement description language, and software (partition and allocation functions) tools. There are three types of allocation models: graph theoretical, integer programming and heuristic.

The article describes TCA problems and contributing factors to thread port-to-port times (tasks in the TCA are arranged into several major processing threads which must meet port-to-port time requirements).

A branch-and-bound heuristic method called TAM (task allocation method) is briefly presented and applied in simulation in an air defense case study.

Mercz, L. I.

HOW TO INCREASE DATABASE PROGRAMMING PRODUCTIVITY.

Infotech 1978, London, 31 pgs.

"If the rate of programmer productivity improvement remains at its current level of 3% per year, by 1985 data processing could be one of the most labour-intensive of all industries.

The aim of this paper is to describe some of the ways in which we could take advantage of the latest evolution of database technology in order to improve programming efficiency over the next five years.

We will investigate only those practical advances of the technology which could be applied right now. Some of these methods can be implemented using currently available software appropriately, with disicpline, and by building higher level canonical functions on today's DBMS.

As the new generation of DBMS will be gradually available, the discipline could be relaxed because the canonical data manipulation functions will be available directly."

Standridge, Charles A.

PERFORMING SIMULATION PROJECTS WITH THE EXTENDED SIMULATION SYSTEM (TESS).

Simulation, vol. 45, no. 6, December 1985, pp. 283-291.

TESS provides a comprehensive, flexible and integrated framework for performing simulation projects. The software capabilites as listed by the author are the following:

1) graphically building SLAM II network and schematic models;
2) forms entry for simulation control information, input and animations;
3) database management of user-defined data and model I/O;
4) preparation of reports and graphs;
5) analysis of simulation results; and
6) animation of runs.

TESS being an "integrated software" allows the user to have minimum knowledge of the operating system. No mention is made of running TESS in heterogeneous systems, nor of capabilities in handling large system simulations.

Standridge, Charles A.

ANIMATED SIMULATION USING *TESS*.

Computer and Industrial Engineering, vol. 10, no. 2, 1986, pp. 121-134.

Animation gives The Extended Simulation System (TESS) the ability to display the dynamics and operating strategies of a simulation graphically. Results from the simulation are stored in the database. Animations are performed as a post-simulation presentation tool using color-raster graphics.

It would be interesting to view results in a network environment where the software resides in a different host computer.

Standridge, Charles A. and Duket, Steven D.

APPLICATIONS OF DATABASE MANAGEMENT AND GRAPHICS IN SIMULATION.

Pritsker and Associates, Inc., West Lafayette, Indiana, 7 pgs.

"As simulation has become a widely accepted analysis technique, the complexity of simulation studies has increased significantly. Consequently, the importance of activities peripheral to model building and implementation has expanded. This increased emphasis on model support activities has led to reliance on database concepts and graphics techniques. This paper describes how these advanced procedures can assist in the simulation modeling process and presents examples of their application."

Standridge, Charles R. and Phillips, James R.

USING *SLAM* AND *SDL* TO ASSESS SPACE SHUTTLE EXPERIMENTS.

Simulation, July 1983, pp. 25-35.

Abstract: New tools simplify the development and simulation of models as well as the analysis and presentation of simulation results. The SLAM language allows the modeler to use network, discrete event and continuous modeling perspectives in any combination to describe the different parts of a system. The SDL database management system allows the modeler to collect simulation results, and at the same time to analyze and present these results independent of the simulation run.

We applied SLAM and SDL in analyzing a package of experiments for the Space Shuttle. The SLAM model of the experimental package shows the utility of a combined network, discrete event, and continuous model. The SDL database illustrates the advantages of storing model run results, using the computer in comparing alternatives, defining and computing performance measures after simulation results are made, and generating reports and plots of model results in the database.

The results of our analysis were also of interest. Only a limited amount of shuttle resources could be used by the package of experiments. Thus, the simulation was used to determine the amount of memory needed for each experiment and for the microprocessor, as well as the reasons for the transmission delay for data from each experiment.

Wiederhold, Gio

VIEWS, OBJECTS, AND DATABASES.

Stanford University, October 1986, 17 pgs.

The major point of the paper is that it appears to be unwise to store persistent objects, which are to be shared, in object format. There appears to be a major conflict between the database

paradigm: making information available for sharing using non-procedural access, and the programming paradigm: devising data structures which are effective representations for procedural access. The author illustrates the problem with an example taken from electronic design applications, where both sharing and object-oriented access is desirable.

A procedural approach is then suggested to permit object-oriented access to use information stored in a relational database. The proposed architecture has the following components:

1) a set of base relations.
2) a set of view-object generators.
3) a set of view-object decomposers and archivers.

An implementation of the sketched approach will have severe performance penalties unless storage transformations are applied to the database and its interface optimized for object-oriented access. However, having a defined interface will permit storage reconfiguration and access optimization, notions inherent in non-procedural data access. A direction for further development, leading towards data independence for object access functions, is indicated.

END

DATE

FILMED

9-88

DTIC